QoS provisioning in converged satellite and terrestrial networks: A Survey of the State-of-the-Art

Christian Niephaus, *Member, IEEE,* Mathias Kretschmer, *Member, IEEE,* and Gheorghita Ghinea, *Member, IEEE,*

***Abstract*—It has been widely acknowledged that future networks need to provide significantly more capacity than nowadays' ones in order to deal with the increasing traffic demands of the users. Particularly in regions where optical fiber are unlikely to be deployed due to economical constraints, this is a huge challenge. One option to address this issue is to complement existing narrow-band terrestrial networks with additional satellite connections. Satellites cover huge areas and recent developments have considerably increased the available capacity, while the cost are decreasing. However, geostationary satellite links have significantly different link characteristics than most terrestrial links, mainly due to the higher signal propagation time, which often renders them not suitable for delay intolerant traffic. This article surveys the current state-of-the-art of satellite and terrestrial network convergence. We mainly focus on scenarios in which satellite networks complement existing terrestrial infrastructures, i.e. parallel satellite and terrestrial links exist, in order to provide high bandwidth connections while ideally achieving a similar end user Quality-of-Experience as in high bandwidth terrestrial networks. Thus, we identify the technical challenges associated with the convergence of satellite and terrestrial networks and analyze the related work. Based on this, we identify four key functional building blocks, which are essential to distribute traffic optimally between the terrestrial and the satellite networks. These are the Traffic Requirement Identification function, the Link Characteristics Identification function as well as the Traffic Engineering function and the Execution function. Afterwards, we survey current network architectures with respect to these key functional building blocks and perform a gap analysis, which shows that all analyzed network architectures require adaptations to effectively support converged satellite and terrestrial networks. Hence, we conclude by formulating several open research questions with respect to satellite and terrestrial network convergence.**

I. INTRODUCTION

Current and emerging networks need to be able to cope with a tremendous increase of traffic volume over the next years [1]. For example, for future 5G mobile networks, which are currently being defined and which are expected to be ready for the market by the end of the decade, a tenfold higher throughput per end user device and thousandfold more traffic in the back-haul segment[1] is anticipated [2]. Similar ambitious goals are now also been set for fixed networks. For example, the European Commission in its Digital Agenda [3] sets the objective to enable broadband Internet connections of at least 30Mbit/s to be available to all EU citizens and 100Mbit/s to at least half of European households by 2020.

This significantly higher amount of traffic volume will pose a major challenge for operators, especially in rural or other difficult-to-serve areas. For example, in the back-haul segment, the deployment of nowadays' typically-used technologies, such as optical fibers, microwave radio links or copper connections [4] is prevented by economical constraints [5]. Moreover, the current approach of overprovisioning the network and purchasing more or enhanced network equipment to ensure congestion free links cannot cope with the expected increase in data traffic. Hence, networks must also be considered as resource constraint, in which bottlenecks might occur [6].

In order to address this issue, a promising approach is to integrate satellite networks as a native component into existing terrestrial infrastructures as already acknowledged previously [7], [8]. Bidirectional satellite networks recently regained the attention of both the scientific and the industry communities since the next generation of Geostationary Earth Orbit (GEO) fixed satellite systems, which are scheduled to be operational by 2020, are targeting the Terabit/s aggregated capacity [9]. These systems will lower the cost per bit significantly mainly by transmitting on different frequencies, i.e. Ka-band, and by using multiple but relatively small spots. Those spots have a size of a few hundred kilometers (instead of e.g a single spot for the whole of Europe) and, thus, allow for more spatial frequency re-use [10]. As satellite links provide ubiquitous and resilient services as well as broadband coverage they are able to deliver high throughput connections and additional capacity wherever it is needed, on a very flexible basis. However, compared to most terrestrial network technologies, both wired and wireless, satellite links have highly different characteristics in terms of latency, burstiness or link stability.

Even though approaches exist that aim at providing triple-play services over broadband GEO satellites, it is de facto

C. Niephaus and G. Ghinea are with the Department of Computer Science, Brunel University, Uxbridge, Middlesex, UB8 3PH, United Kingdom. (e-mail: christian.niephaus@brunel.ac.uk, george.ghinea@brunel.ac.uk

M. Kretschmer is with the Fraunhofer Institute for Applied Information Technology (FIT), Schloss Birlinghoven, 53754 Sankt Augustin, Germany. (e-mail: mathias.kretschmer@fit.fraunhofer.de)

This work has been supported by the BATS research project which is funded by the European Union Seventh Framework Programme under contract n317533. The views and conclusions contained here are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the BATS project or the European Commission.

[1] We refer to back-haul network as the network connecting the access segment, i.e. last-mile, with the core network, which is often also referred to as transport network or aggregation network.

impossible to achieve a similar Quality-of-Experience (QoE) to terrestrial networks, when real-time and interactive application are being used [11]–[13]. The higher latency on satellite connections as a result of the high signal propagation time inevitably means the user's QoE is lower.

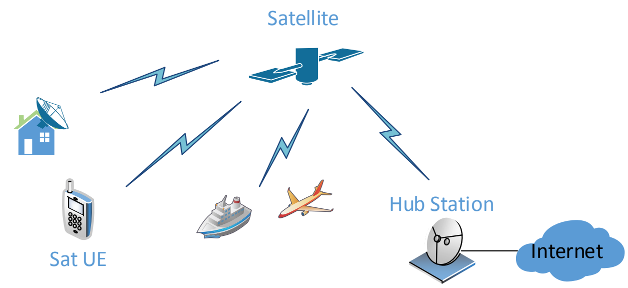
On the other hand, satellite networks are primed for broadcast and multicast type services, since they can reach a potentially unlimited amount of receivers with a single transmission. Given the predominance of video traffic in future networks [14] as well as emerging edge caching approaches, such as [15], satellite networks are seen as an option to tremendously lower the load on the terrestrial back-haul networks, since both services can highly benefit from broadcast distribution. Moreover, ongoing trends of broadcast and broadband convergence to allow hybrid TV applications [16], [17] are also fostered by a close integration of satellite networks and existing terrestrial networks.

In this paper we analyze the current state of the art of converged satellite and terrestrial networks used to provide ubiquitous connectivity in rural and remote areas. It should be noted that we assume multi-beam bi-directional GEO satellite connections, e.g. based on Digital Video Broadcasting-Satellite - Second Generation (DVB-S2) [18] and DVB Return Channel via Satellite (DVB-RCS) [19]. Moreover, we also assume Fixed Satellite Systems (FSS), even though several aspects discussed in this work are also applicable for Mobile Satellite Services (MSS).

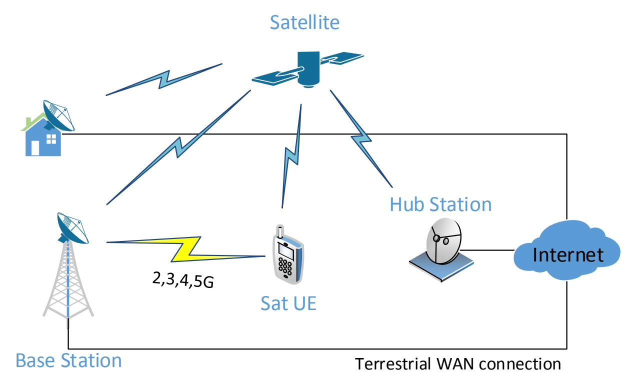
The remainder of the paper is structured as follows. In section II, we describe concrete convergence scenarios considered in this work. In section IV, we elaborate on the technical challenges associated with convergence of satellite and terrestrial networks, while in section III we analyze the existing literature and then in section V, we identify key functional building blocks for converged satellite and terrestrial networks, based on previous analysis of the challenges. In section VI we evaluate current state-of-the-art network architectures with respect to these key functional building blocks and analyze their gaps in section VII. In section VIII, relevant Key Performance Indicators (KPIs) are discussed that allow for assessment of different converged satellite and terrestrial network approaches. In section IX, we summarize the open research items before we conclude the paper in section X.

II. Convergence Scenarios

Satellite networks are jointly used in various scenarios with terrestrial networks. Historically satellite networks are often used to extend terrestrial networks in order to provide connectivity to areas without any terrestrial connection, as shown in Fig. 1a. These scenarios typically include the connection of rural and remote households and premises or moving locations, such as a vessel or an airplane [20]. Typically on one or both edges of the satellite link, terrestrial networks are connected. For instance, in a remote household typical end user devices are connected locally via a Small Office/Home Office (SOHO) (W)LAN router, which establishes the satellite connection via a satellite modem. However, it is important to note that the



(a) Satellite network extending a terrestrial network



(b) Satellite network supplements a terrestrial network

Fig. 1. Comparison of satellite link integration options

satellite link is the sole connection between the terrestrial edges of the network, so that the traffic cannot be routed differently. Such a setup usually aims at providing general connectivity.

Various scenarios and use cases following this setup have already been widely discussed and surveyed in the literature, e.g. [10], [20]–[25]. Hence, in this work we focus on scenarios where the satellite network provides an additional connection in parallel to an existing terrestrial one, as shown in Fig. 1b. A high capacity satellite network supplements an existing terrestrial connection in order to increase the performance of a terrestrial network. Such a scenario makes sense if the terrestrial link performs insufficiently and is not able to cope with the traffic demands. In contrast to the previous setup, the satellite link provides not the sole connection but an alternative one. Thus, traffic needs to be distributed properly onto the terrestrial and the satellite network.

In the following we present two concrete exemplary scenarios, which benefit from an additional satellite link. The first is a remote household, connected with a poor last mile connection, such as a slow X-Digital Subscriber Line (xDSL) connection. The second scenario encompasses remote mobile Base Stations (BSs), i.e. Evolved Node Bs (eNodeBs), connecting typical Mobile Terminals (MTs) as end user devices.

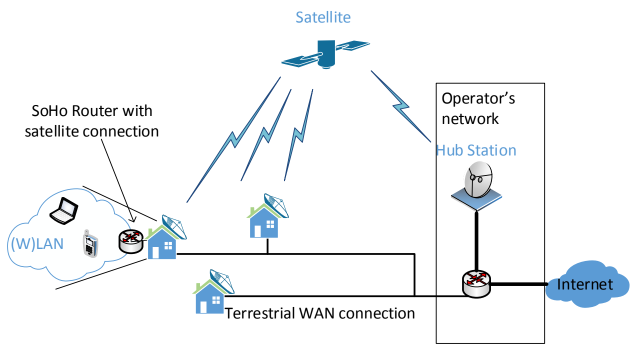


Fig. 2. Overview remote household scenario

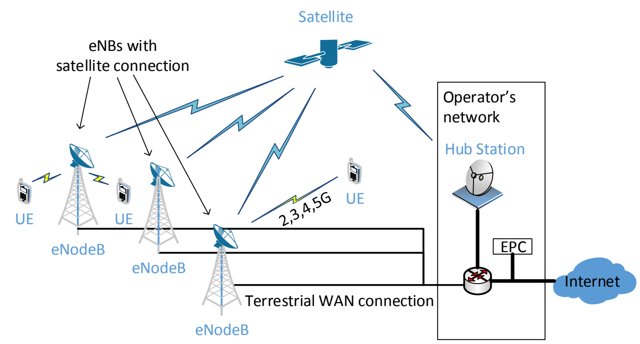


Fig. 3. Overview remote mobile BSs scenario

A. Remote household

Many areas all over the world, particularly in rural and remote regions, lack terrestrial high-speed broadband connections, which are able to cope with future traffic demands. Moreover, a low population density and large areas often render upgrading existing narrow-band terrestrial connections economically unappealing. In this case, the fast and flexible deployment characteristics of satellite connections can be exploited to establish additional satellite links, which do not replace but rather supplement the terrestrial connection [26], as shown in Fig. 2. It should be noted that this is different to setups where the satellite connection is only used for traffic towards the end user, which is referred to as downlink, and the terrestrial connection for traffic towards the operator's network, i.e. the uplink. These setups, which have been commercially available, only require a unidirectional satellite connection and distribute the traffic statically based on its direction. In contrast, the scenario we are considering utilizes both terrestrial and satellite connections, simultaneously in both traffic directions.

Individual houses and premises are equipped with an additional satellite modem in order to increase the available overall capacity. Both the terrestrial Wide Area Network (WAN) as well as the satellite connection terminate at a typical SOHO router connected to or equipped with a satellite modem. End user devices are connected to this router using typical (W)LAN connections, such as Ethernet or IEEE 802.11 WLAN. On the other edge of the network on the operator site a corresponding core router establishes a link to the satellite hub. Thus, the additional satellite connection can be used to either increase the resilience or to offload traffic from from the terrestrial networks in order to overcome bottlenecks [27]. However, if both connections are used simultaneously it needs to be decided which type of traffic is sent over which link. This applies to both the uplink as well as the downlink.

A cost benefit analysis conducted by the EU Broadband Access via integrated Terrestrial and Satellite systems (BATS) research project [28] found that such a scenario can help to reduce the cost for broadband penetration in rural and remote areas. According to this study, an investment into satellite technologies to deliver 30Mbit/s also in rural and remote areas is up to 25% cheaper per household than the incremental investment required to upgrade fixed line infrastructures to achieve the same bandwidth . Moreover, the findings of [28] suggested that such a converged satellite and terrestrial scenario can play a role if the terrestrial fixed line speed for a household is between 0-8Mbit/s.

B. Remote mobile base station

Just like remote houses, mobile BSs, particularly in rural and remote areas, can also be equipped with an additional satellite connection, as depicted in Fig. 3, in order to increase the performance of the back-haul connection.

Future networks are expected to not only provide best-effort (BE) services. Instead, guaranteed services need to also be supported in order to enable novel and emerging applications. Examples of these applications include high definition video streaming, cloud-based applications, web conferencing or even Machine-Type-Communication (MTC), all of which have different requirements in terms of latency, required bandwidth, jitter and reliability [29] that need to be considered in order to satisfy the user's demands and achieve a high QoE [2].

Given that, it is beneficial to offload certain traffic to the satellite network if the terrestrial back-haul connection does not provide a sufficient performance. However, the additional back-haul connection via the satellite needs to be transparent to the end users and their MTs, so that no changes on these devices are necessary.

Furthermore, selecting the optimal connection for a particular portion of the traffic is a highly crucial decision that needs to be made. Due to the different and more varying link characteristics of satellite networks only certain traffic types are suitable to be sent via a satellite link in order to allow for a good QoE.

Moreover, if multiple receivers in an area consume the same content, this scenario can highly benefit from the satellite's broad- and multicasting capabilities. For example, high bandwidth video traffic can be effectively offloaded to the satellite network and distributed with a single transmission to multiple receivers. Particularly when used jointly with intelligent caching mechanisms, this unique satellite characteristic is significantly helpful to reduce the network load.

III. Review of related work

Satellite networks have been considered widely in the literature for many years. Various aspects related to the convergence of satellite and terrestrial network have therefore already been discussed. In the following we summarize and classify the existing literature in order to identify the gaps with respect to our exemplary scenarios.

A. Quality-of-Service (QoS) in satellite networks

One category of publications deals with the implementation of QoS in satellite networks. More than a decade ago, a holistic overview on QoS over satellite networks was presented in [21]. The authors survey a satellite network architecture and analyze relevant QoS mechanisms in different layers, starting from the physical layer and the impact of adaptive coding and modulation (ACM), via the link layer and the Demand Assigned Multiple Access (DAMA) process, which allocates resources to individual terminals (see section IV-A2), through to the interaction with Internet Protocol (IP) QoS mechanisms such as Differentiated Services (Diffserv) and Integrated services (IntServ). Back then, [21] identified several open issues in order to allow for QoS guarantees in satellite networks, including bandwidth allocation and Transmission Control Protocol (TCP) enhancements for satellites.

The open issues mentioned by [21], are partially addressed by [23] and [30]. Both present a QoS architecture for satellite links based on DVB-RCS and deal with the mapping of IP layer QoS mechanisms onto DVB-RCS. This includes in particular how capacity for different traffic classes is assigned by the DAMA process. Furthermore, [30] proposes a detailed architecture to provide end-to-end QoS over satellite networks. The authors rely on the Session Initiation Protocol (SIP) protocol to allow applications to signal their QoS requirements. For applications not using SIP protocols a QoS agent running on the end user device is proposed, which allows an application to announce its QoS requirements to the system.

In contrast to the generic QoS architectures presented by [23] and [30], several publications focus on enabling specific services over satellite links. For example, [31] analyzes the impact of different dynamic bandwidth allocation schemes for the satellite uplink on Voice over IP (VoIP) connections. They conclude that statically assigned capacity is not optimal in a satellite environment as it wastes resources and therefore becomes inefficient. Moreover, according to their results, even with dynamic resource allocation a similar quality of the voice call is possible. It should be noted, though, that the authors do not compare their performance results against terrestrial VoIP connections, which are not impacted by the high satellite fixed latency.

In related work [20] Medium Earth Orbit (MEO) satellite networks as a back-haul network for BSs, i.e. eNodeBs, are considered, particularly for moving BSs on vessels, airplanes or buses. The authors focus on quality of video services, including Video on Demand (VoD) and video communication. They run a set of simulations, which show that the quality for VoD is good if the network is not heavily loaded, while video communication only works in lightly loaded networks. However, in highly loaded networks VoD suffers from TCP problems. It should also be noted that the authors advocate a generic interface for adding meta-data, such as priority or tolerable latency, to the payload data transmitted over the satellite.

B. Protocol adaption for satellite links

Another category of publications deals with the adaptation of protocols to make them better suited or even functional for satellite links. As far back as in 1997 [25] presented a satellite Asynchronous transfer mode (ATM) architecture, which aimed at using ATM networking technology via satellite networks. The authors concluded that the use of ATM over satellite links requires major modification on typical control mechanisms, such as congestion and traffic control, as well as optimization on the satellite link, e.g. the bandwidth management.

More recently, [32] proposed a Long Term Evolution (LTE)-satellite component. Their focus is to adapt LTE mechanisms and algorithms to allow for using LTE over satellite links with their high latency. This includes modification on basically all layers. The authors introduced a new radio interface and a new virtual Hybrid automatic repeat request (HARQ) scheme, since the typical LTE HARQ is designed for an 8ms round-trip time (RTT). Moreover, handover and paging mechanisms of LTE are also required to be updated in order to use LTE over satellite.

C. Satellite performance optimization

Several publications exist that focus on enhancing and optimizing the performance of the satellite links so that they better fit into a terrestrial infrastructure. One example is [33], which proposes a Delay-tolerant networking (DTN) [34] approach to deal with the long latencies on satellite networks in order to improve the performance of LTE voice calls. By introducing an additional layer between the application and transport layer, which provides store-and-forward message capabilities, the authors' proposal can, transparently to the application layer, split the end-to-end TCP connection. This way, similar to Performance Enhancement Proxies (PEPs) (see IV-A1), different transport layer protocols can be used on the satellite link. The authors' results show a performance improvement of their approach compared to regular TCP protocols in terms of throughput.

Similarly, [35] implemented a PEP solution and compared this approach against different TCP variants with respect to performance and reliability. Results show that the authors' PEP solution can mitigate the latency drawback of satellites for TCP connections in terms of throughput. Depending on the actual RTT, the approach can achieve up to ten-times higher throughput on the application layer.

Optimizations on the physical layer are discussed by [36]. More precisely, the authors analyzed the potential of multiple-input and multiple-output (MIMO) transmission on satellite links for fixed and mobile satellite systems. While satellite systems can profit from MIMO techniques the authors identified a couple of open research items and inconsistencies with the current DVB-S2 standards, in particular with respect to ACM.

D. Handover and mobility management

Another category of existing publications deals with handover and mobility management in integrated satellite and terrestrial networks. This becomes important once MSS are considered. In such networks different handovers need to be distinguished, namely horizontal and vertical handovers. Horizontal handovers in satellite networks are needed when a satellite MT moves from one spot-beam to another. In contrast, when a satellite MT hands over from a satellite network to a terrestrial mobile network or vice versa, this is considered to be a vertical handover.

A general discussion and comparison on mobility management solutions that allow for global roaming in heterogeneous (wireless) networks can be found in [37]. The authors proposed a novel mobility management architecture relying on so-called interworking agents. Such a network interworking agent differentiates and supports intradomain (horizontal) and interdomain (vertical) mobility. While for the former existing protocols are used, a cross-layer management approach is proposed for the latter. This cross-layer approach aims at detecting early possible interdomain handoff destinations to allow for authentication, authorization and mobile IP registration prior to the actual handover.

An approach for link selection in integrated satellite and terrestrial networks is also presented in [38]. The authors focus on a handover mechanism, which relies on the IEEE 802.21 standard [39] and integrate it into a satellite architecture. They aim at selecting the most suitable link based on the current link conditions.

It should be noted, however, that in both publications the handover is always performed on a per device basis. That is, the whole traffic to and from a device is routed either via the satellite link or terrestrially. However, just a single flow cannot be handed over to a particular network, e.g. while a video is streamed via the satellite, the remaining traffic cannot be sent through the terrestrial link.

E. Spectrum sharing between terrestrial and satellite networks

Various papers analyzed if and how radio frequencies can be shared between satellite and wireless terrestrial networks. As we explain in section IV-A4, wireless terrestrial and satellite networks might use the same frequencies in order to allow for an increased spectrum efficiency, which on the other hand might lead to interference.

For example, [40] investigates by simulation the co-frequency interferences if mobile networks and mobile satellite networks operate in the same spectrum. The authors proposed a methodology to evaluate this interference. They conclude that, particularly in the uplink, interferences may occur heavily if both networks operate on the same frequency in the same geographical area. Also [41] analyzes an integrated bi-directional satellite system consisting of a satellite and a terrestrial network, which both operate on the same frequency band. The authors evaluate the interferences between both systems and the impact of so-called exclusion zones. In an exclusion zone the satellite antenna gain is below a given isolation value from the center of the beam. They conclude that a few terrestrial users are responsible for most of the interferences.

F. Integration of satellite and terrestrial networks

The idea of integrating satellite into terrestrial networks has also been present in articles for several years. In 2005 [8] present business and market observations for satellite networks in general. The conclusion is that integration of satellite networks and terrestrial networks is the key to success for satellite networks and that there is a mutual benefit for both the terrestrial and satellite world. Unfortunately, the authors do not present concrete technical solutions.

Similarly, [42] deals with scenarios of converged satellite and terrestrial networks. The authors identify associated issues, which include transmission efficiency, resource allocation and mobility management. Therefore they present enhanced TCP methods as well as a cross-layer bandwidth allocation approach, which address the issue of varying physical layer capacity on satellite links caused by ACM.

An overview of issues related to integrated and hybrid satellite and terrestrial networks is provided in [43]. The authors consider integrated networks, when satellite and terrestrial networks are operating on the same frequency, while hybrid satellite/terrestrial networks are interconnected but operate independently. The paper gives an overview of typical issues in such systems. These include physical layer issues, such as MIMO, but also resource management, handover issues or QoS problems. Finally, Traffic Engineering (TE) concepts are being discussed, which show that the satellite network can reduce the blocking on terrestrial links. However, in this regard the focus is on bandwidth management, without considering other QoS requirements of the traffic.

Finally, [17] discusses the advantages of the convergence of broadcast and unicast (cellular) networks. Given that a lot of content requested by users in a network is actually the same, e.g. top10 YouTube videos, the authors claim that it is often advisable to broadcast the content rather than transmitting it individually. This way, the load on unicast transmission networks can be reduced. However, in order to allow for broadcasting, the content needs to be requested relatively simultaneously, and [17] identified technical challenges in this regard, which have not yet been solved. Firstly, correlated content that can actually be broadcasted needs to be detected. Secondly, situations when it is worth broadcasting content, i.e. the number of receivers is high enough, must be identified, and, finally, the content transmission needs to be synchronized.

G. Conclusion of existing literature overview

In Table I we summarize the existing work and their key points. As can be seen, many existing publications address typical challenges that occur when satellite networks are used, regardless of the concrete integration with terrestrial infrastructures. We identify three groups of existing work for which this is applicable: Firstly, publications that optimize the satellite performance as such on various layers. For instance, [36] addresses optimization on the physical layer while [35] and [33] aim at transport layer enhancements. Secondly, different

Table I
OVERVIEW OF EXISTING WORK

Reference	Key points
QoS in satellite networks	
[21]	Holistic survey of QoS mechanisms on satellite networks on various layers, i.e. physical, data link, network and transport layer.
[30], [23]	Presentation of a QoS architecture for DVB-RCS that allows for providing triple-play services over a satellite link and proposal of mechanisms how typical IP QoS mechanisms can be mapped onto a DVB-S/DVB-RCS architecture.
[31]	Analysis of the impact of different bandwidth allocation schemes for VoIP connections over satellite links.
[20]	Analysis of the suitability of MEO satellite for LTE back-hauling, particularly for VoD services.
Protocol adaption	
[25]	Discussion on issues related to the ATM protocol and its implementation over satellite networks.
[32]	Proposal of a LTE-satellite component, which modifies LTE mechanisms and algorithms in order to adapt those to the satellite radio transmission environment.
Satellite performance optimization	
[33]	Integration of a DTN architecture into a satellite-LTE network in order to increase the TCP performance over satellite links.
[35]	Presentation of a PEP solution and comparison of it against various TCP variants.
[36]	Discussion on physical layer optimizations and survey of the potential of MIMO techniques for fixed and mobile satellite systems.
Handover and mobility management	
[37]	Survey of mobility management solutions for roaming in various heterogeneous networks including a satellite.
[38]	Proposal of a handover solution based on IEEE 802.21 that aims at selecting a terrestrial or a satellite link based on the current link conditions.
Spectrum sharing	
[40], [41]	Analyses of co-frequency interferences between mobile and satellite networks when operated on the same frequencies.
Integration of satellite and terrestrial networks	
[8]	Discussion on business and market opportunities for satellite networks in general and integrated satellite/terrestrial networks in particular.
[42], [43]	Analysis and survey of issues with respect to interworking of satellite and terrestrial networks on various layers.
[17]	Survey of the potential and the challenges of converging broadcast and cellular networks.

authors analyzed specific network protocols with respect to their applicability to be used over satellite links. [25], for instance, focused on ATM communication over satellite links while [32] proposed adaptations necessary to operate LTE on satellite connections. Finally, the implementation of QoS in integrated satellite networks is often considered, e.g. [23] and [30] proposed a QoS architecture that allows for prioritizing traffic and a proper mapping of IP and Digital Video Broadcast (DVB) QoS mechanisms.

However, as we explained in section II, we consider a specific integration scenario, in which parallel satellite and terrestrial paths exist. That is, we acknowledge the relevance of these publications for all networks involving a satellite, yet we believe that the scenarios we are dealing with have additional challenges that are have not been addressed in the existing literature. As we elaborate further in section IV, in case of parallel satellite and terrestrial connections, traffic needs to be distributed onto both links based on its requirements and the links' capabilities. This aspect is missing in the aforementioned publications.

Similarly, several publications focus on spectrum sharing aspects between satellite and wireless terrestrial networks, such as [40] and [41]. Those aspects are mainly relevant for scenarios with moving MTs that are able to connect to both satellite and wireless terrestrial networks.

Furthermore, several publications discuss different handover scenarios. Examples of such are [37] and [38]. Particularly the latter deals with selecting a terrestrial or satellite link. However, the link selection is typically being performed on a device basis. That is, all traffic to or from the MT is routed over the same link. The granularity of link-selection required in the considered scenarios should be more fine-grained, so that two traffic flows with different requirements can be routed via different links.

Finally, various publications exist, which particularly target converged satellite and terrestrial networks. While [8] argues for the convergence of both networks, the authors lack concrete technical solutions; [42] however suggested some approaches. Unfortunately, [42] also does not focus on the problem of distributing traffic between the terrestrial and the satellite network, yet the authors see the need for a ”context-aware routing schema” in order to enhance the end-to-end QoS.

IV. TECHNICAL CHALLENGES OF CONVERGED SATELLITE AND TERRESTRIAL NETWORKS

In the following we present the main challenges associated with the convergence of satellite and terrestrial networks. We differentiate between two categories of technical challenges: The first group arises from the usage of a satellite link within a network in general, while the second group applies to scenarios in which satellite links establish an alternative connection, in parallel to an existing terrestrial one, as described in II-A and II-B.

A. Satellite-inherent challenges

The satellite-inherent challenges that need to be dealt with include differences in the high latencies on satellite connections, frequently varying link conditions, the medium access coordination for the uplink as well as overlapping radio frequencies with terrestrial wireless networks. Those challenges apply to all satellite networks.

1) High latency on satellite connections: A major difference between satellite and terrestrial systems is the latency. More precisely, GEO satellite links have a high fixed latency, as depicted in Fig. 4. While dynamic latency consists of the time required to serialize data, process a packet at a network entity as well as potential queuing and buffering delays and

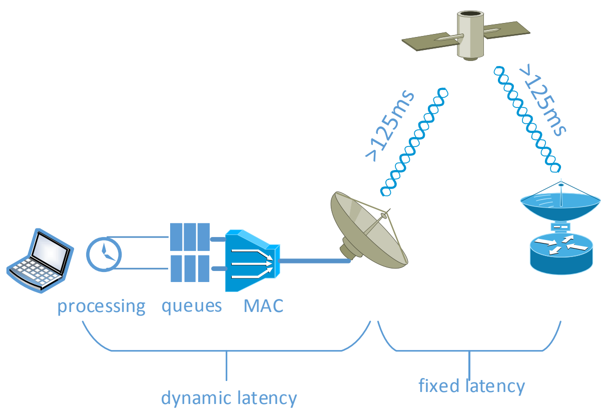


Fig. 4. Latency model of a GEO satellite link

the time required to get access to the medium, the fixed part of the overall latency is the actual signal propagation time that a packet experiences in any case when being transported over a network medium. It is a physical characteristic, mainly restricted by the speed of light. Moreover, the dynamic latency highly depends on the available capacity on the link and can be reduced and controlled by prioritization, packet dropping or admission control to avoid congested links.

GEO satellites operate at a height of approximately 36.000km. Hence, the fixed latency on a satellite link is in the order of magnitude of a few hundred milliseconds. While the high capacity of satellite links can be a benefit for exchanging large amounts of data, the additional delay has a negative impact on the QoS and on applications provided via satellite that are not tolerant against high latencies. An example of these are real-time or interactive applications. Furthermore, protocol state machines and specific functionalities may also need to be adopted to cope with the higher latency. For example, the (standard) congestion control mechanisms of the TCP [44] rely on the RTT and thus the TCP throughput suffers significantly from the high latency on the satellite links.

In order to mitigate this, specialized TCP algorithms, such as [45], or TCP-PEPs [46] are commonly used. PEPs are TCP proxies, which break the TCP end-to-end connection at the satellite modem and/or the hub station, so that the satellite portion of the end-to-end path is separated from the rest. That is, instead of a single end-to-end TCP connection up to three intermediate connections are established, namely between the end host initiating the connection to the first edge of the satellite link, between the sat modem and the hub, and finally between the other edge of the satellite link the and the destination host. On the one hand, this allows for running an adapted TCP protocol or a completely different transport layer protocol optimized for the high latency, but on the other hand, it breaks the end-to-end paradigm of TCP and leads to issues with layer 3 encryption mechanisms [35].

2) Medium access coordination and resource reservation: Another major challenge in this regard is to coordinate the uplink connections (towards the core network) in order to provision QoS. In satellite networks on the downlink only a single sender, namely the hub station, exists, while on the uplink potentially many highly distributed senders access the shared medium. In order to do this, the DVB Return Channel via Satellite - Second Generation (DVB-RCS2) standard [47] specifies transmission-slots during which they can transmit. These are assigned to each sender by a central entity called network control center (NCC) located at the hub station. The NCC periodically broadcast the assignment of transmission-slots, which are either based on Service Level Agreements (SLAs) and constantly assigned, or on dynamic requests from a sender depending on the current traffic demands. The latter is called DAMA. Depending on the traffic, a satellite terminal can request different categories of capacity. For services requiring a higher priority or real-time services typically constant rate assignment (CRA) or rate-based dynamic capacity (RBDC) are used, which provide capacity guarantees. In contrast, for lower priority and BE traffic, volume-based dynamic capacity (VBDC) and free capacity allocation (FCA) capacities are requested. These provide fewer guarantees. The NCC grants permissions to transmit by sending a terminal burst time plan (TBTP). Unfortunately, the standard lacks a definition on how the capacity requests are triggered. Since those requests and the allocation responses traverse the satellite link as well as the actual data, the overall latency increases even further. Thus, higher priority, real-time and more interactive traffic should be mapped onto constantly assigned slots. However, capacity, which is fixed and assigned to a station, is wasted if the station has nothing to send in a particular slot. Hence, there is a trade-off between the fixed allocated capacity to certain stations and the capacity dynamically allocated.

3) Changing link conditions: The quality of the satellite link is affected by weather, such as heavy rain, much in the same way as many other technologies transmitting wirelessly, and therefore changes more frequently. In order to avoid wasting spectral efficiency due to significant Signal-to-Noise Ratio (SNR) buffers, satellite connections implement an ACM mechanism. ACM adapts to these changing conditions by modifying the used Modulation and Coding Scheme (MCS). That is, if e.g. heavy rain appears, the modulation is changed to a more robust schema, based on feedback the receiver sends to the sender. On the other hand, the increased robustness comes with the additional cost of reduced capacity, as more redundancy is added to the signal. Once the rain stops, a more efficient MCS can be selected again. Thus, the link capacity that can be used by upper layers changes too.

Furthermore, given that ACM is designed to quickly adapt to changing wireless channel conditions, these changes might occur frequently without prior notification. Hence, capacity planning and traffic shaping become significantly more difficult [48].

4) Overlapping frequencies: As radio spectrum has increasingly become scare, resource competition between terrestrial wireless networks and satellite networks can occur. Satellite and wireless terrestrial networks, e.g. cellular networks, can either share the same frequencies or operate on a dedicated spectrum. Operating on the same frequencies allows for reuse of spectrum and thus for increasing the spectral efficiency.

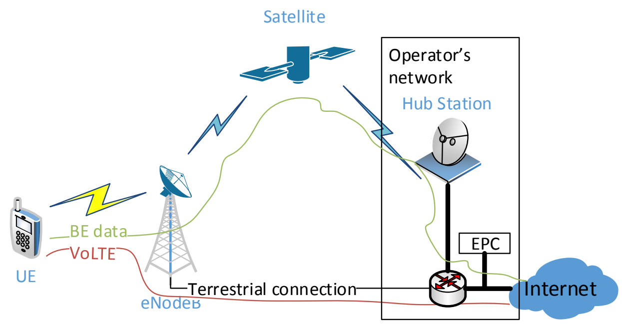


Fig. 5. Data paths in a integrated satellite/terrestrial network

This frequency reuse is particularly interesting for moving MTs, which are able to connect to both wireless terrestrial and satellite networks. However, if not managed properly, both parts might interfere with each other. Hence, it is usually required that both networks are managed by the same operator, if the available spectrum is to be shared, as tight coordination is required.

B. *Specific challenges for parallel satellite and terrestrial link scenarios*

When a satellite network supplements terrestrial infrastructures, the network topology becomes more complex than e.g. a tree, since alternative parallel paths are available between two nodes in the network. However, unlike parallel links in a plain homogeneous network, the different links in a converged satellite and terrestrial environment differ extremely in their characteristics. Hence, the major challenge is to decide whether a particular chunk of traffic should be routed via the satellite or terrestrially.

An example is depicted in Fig. 5. Two different paths exist between the eNodeB and the core network. Hence, for the uplink traffic what needs to be decided is if the satellite or the terrestrial link should be used at the eNodeB and the same decision needs to be made in the operator's network for the downlink traffic. As implied in this simplistic figure, ideally the path selection is not static or simply based on source or destination IP addresses, but rather dynamic and on a fine-granular basis, since traffic of different services exchanged between the same source and destination might need to be routed differently. That is, the selected path should depend on the one hand on the network capabilities and on the other on the requirements the traffic imposes on the used network in order to allow for a high QoE. Thus, traffic with the same source and destination IP address might take different routes depending on its requirements. Even though at first glance this might seem trivial given the amount of matured IP routing protocols, the unique characteristics of satellite links tremendously increase the complexity. For example, given the characteristics of a GEO satellite link, a real-time and interactive application, e.g. VoIP, will rather use the terrestrial link due to the latency aspects, whereas latency tolerant applications, such as a file transfer, might prefer the satellite if sufficient capacity is available.

Therefore, TE, which is the process of "enhancing the performance of an operational network, at both the traffic and resource level" by "addressing traffic oriented performance requirements, while utilizing network resources economically and reliably" [49], becomes essential. More precisely, satellite specific-aware mechanisms are required, which allow for this fine-grained kind of traffic steering implied in the previous example, so that both terrestrial and satellite networks can benefit from the mutual integration. Unfortunately, TE in typical terrestrial networks focuses mainly on putting "the data traffic where the network bandwidth is available" [50], [51], which is not sufficient in converged satellite and terrestrial networks due to the different characteristics of satellite links, in particular the high fixed latency. Instead of focusing solely on managing the available capacities, additional parameters are required to be taken into account, such as the tolerable latency, the advantages of the satellite with respect to broadcast traffic or the impact of varying link conditions on the overall QoE. Hence, we identified three main challenges associated with converged satellite/terrestrial network scenarios, which encompass parallel links, namely traffic requirement and link characteristics identification as well as path selection. In the following we provide a detailed discussion on these challenges.

1) Traffic requirement identification: To guarantee a good service quality and eventually a good QoE perceived by the end users in converged satellite and terrestrial networks, as an absolute minimum it needs to be known if the traffic is suitable to be routed via the satellite in terms of its latency requirements. If, for example, a real-time service requires a guaranteed latency below 250ms in order to achieve a good QoE, the service might not be usable over a GEO satellite connection, regardless of the available capacity on that link. This is a major difference compared to pure terrestrial networks, where capacity management and congestion avoidance might be sufficient, without the need for determining the actual traffic QoS requirements, since high latencies or jitter values as well as packet losses can often be avoided by preventing a congestion in the network.

It should be noted, however, that this becomes increasingly important in today's operator networks independent of the satellite network integration, since many application providers require special services for their applications, particularly rich media and pervasive video applications, which have enormous QoE expectations as revealed by recent Internet Engineering Task Force (IETF) activities [52]. Moreover, in order to perform adequate TE, knowledge of traffic requirements, such as required bandwidth or latency demands is also worthwhile.

However, becoming aware of the these is not a trivial task. Various methods are available to identify applications or application types just by monitoring the traffic. The examination of the transport-layer port numbers, which has been the preferred method for several years, has proven to be inaccurate in nowadays' networks [53], e.g. due to re-usage of well-known ports or higher layer tunneling. This leads to more complex methods, such as Deep packet inspection (DPI) or machine learning approaches [54]. However, encryption, frequently emerging novel applications, or privacy and scalability issues lead to inaccuracies with those approaches as well [55]. Furthermore, the application identification mechanisms are usually proprietary, with non-standardized interfaces [56]. That is, there is neither a standardized interface to retrieve information from a traffic requirement identification process, nor a defined set of identified parameters. For example, a DPI box from vendor A might provide different information than a box from vendor B using a completely different interface to access the information. Moreover, even if the application or the application type is known, a mapping on the concrete QoS requirements needs to be defined.

An alternative to identify IP traffic by monitoring and analyzing it is the usage of meta-information sent by the application or application-related protocols. Those meta-information could contain application type or even the traffic QoS requirements of the application. The Resource Reservation Protocol (RSVP) [57] or Next Steps in Signaling (NSIS) [58] aim at reserving resources end-to-end and thus allows for signaling information related to the traffic. For example, the RSVP SENDER_TSPEC object [59] allows a sender to specify the generated traffic by providing a peak data rate as well as a token bucket rate and a token bucket rate. However, those approaches are not widely implemented in applications due to their overhead and limited deployments [56]. Proxy extensions for RSVP have been defined in [60], [61] to allow for the usage of RSVP if the application does not support it. However, the challenge is then to make the sender proxy aware of the application requirements. Different methods to do this are suggested in [60]. Most of them rely on application layer protocols which also need to be supported by the application. Application layer protocols, such as the SIP [62] and Session Description Protocol (SDP) [63], are often used to signal and control multimedia communication. SDP aims at conveying meta information on the actual communication, including the type of media, i.e. audio, video, etc., and the required bandwidth. However, these protocols are only in use for a very limited amount of applications, leading to current discussions of more lightweight and generalized mechanisms in the standardization bodies [64].

Furthermore, given the broadcast capabilities of the satellite network, it is highly beneficial to identify traffic worth broadcasting due to a high number of recipients. Exploiting the unique broad- and multicast characteristics of satellite networks to reach a high number of receivers with a single transmission can significantly offload traffic from terrestrial networks [17].

Hence, the first challenge that needs to be solved is to determine the traffic's QoS requirements, either by performing reliable packet analysis or by allowing applications to send them upfront with a reasonable overhead, and to provide this information in a well-defined form, so that it can be used by other protocols and algorithms.

2) Link characteristics identification: Knowing the requirements of the traffic flows is only one aspect. In order to effectively use this information to control the traffic flow, awareness of the link (and eventually the path) characteristics is essential to match the traffic's requirements against the network capabilities. For example, in order to decide that the VoIP traffic depicted in Fig. 5 is better sent via the terrestrial link, the entity selecting the link needs to know that the satellite link's latency exceeds the maximum tolerable latency for VoIP communication with a good QoE.

It is important to note that, due to the usage of ACM, the characteristics of satellite links might change over time. As described in section IV-A3, the capacity on a link might increase or decrease depending on the current MCS selected by the ACM mechanism. This is considerably different compared to wired connections where the link characteristics are rather stable as long as the links are not congested.

Hence, it is not sufficient to configure the link characteristics manually or to identify them once the link is established. Instead, it has to be a continuous process, so that an entity, which selects the proper link, can consider this in its decision process.

Existing approaches to identify the link or path characteristics can be generally classified in three categories:

- Transport layer (L4) mechanisms
- Active measurements
- Lower layer (L1 and L2) mechanisms

Methods operating on the transport layer aim at determining the available bandwidth by monitoring TCP parameters, such as the congestion window. For example, [65] designed a new TCP scheme especially suited for wireless networks that consists of an available bandwidth estimation algorithm. Based on the amount of in-flight data, RTT as well as received Acknowledgments (ACKs) rate and the amount of acknowledged data, the current bandwidth is estimated. Simulations show that the available bandwidth can be properly determined. However, the frequent usage of PEPs in satellite networks, as described in section IV-A1, which breaks the end-to-end paradigm of TCP, renders a complicated environment for such a TCP-based link characteristics estimation approach. Instead of having a single end-to-end TCP connection, up to three TCP connections are established or even a different transport layer protocol is used on the satellite link. Furthermore, the available bandwidth estimation in [65] determines the available bandwidth for TCP but does not work for different transport layer protocols, such as User Datagram Protocol (UDP).

Furthermore, approaches are available, which estimate link or path characteristics by actively creating and measuring traffic, such as [66] that estimates the available bandwidth on an end-to-end path by measuring the one way delay of artificially created periodic streams. The author's idea is that the latencies of a periodic flow show an increasing trend if the rate is greater than the available bandwidth. It is shown in [66] through simulations and Internet experiments that this approach can measure the available bandwidth accurately under various load conditions. Moreover, even though additional network load is introduced by sending measurement probes, the approach is non-intrusive, since each measurement stream only contains 100 packets. It should be noted, however, that the experiments are done in wired networks. The more volatile conditions on satellite links as well as the impact of the dynamic capacity allocation on the satellite's return channel are additional challenges, which are not considered by the authors.

Finally, approaches exist that exploit link layer (L2) information to improve the accuracy of the link characteristic identification. For example, the satellite's link layer is aware of currently used MCS, the amount of assigned capacity or the fill level of queues, so that the available bandwidth or latency can more easily be determined. However, while the first two methods can work without further knowledge of the underlying technology, the last approach requires awareness and assistance of the link layer technology. A couple of scenarios, in which a system could highly benefit from information available at L2 in order to improve the overall performance, have already been identified in [67]. One of these is traffic shaping on links where the capacity changes frequently over time and a static or manual configuration is not appropriate. This approach is supported by IETF activities to standardize the Dynamic Link Exchange Protocol (DLEP), which aims at allowing a routing entity (L3) to retrieve information on link status and the link capabilities from a modem in a technology-independent manner [68]. This protocol is designed to operate between a router and its attached modems, particularly in heterogeneous environments, allowing the modem to propagate the link capabilities and other relevant events, e.g. link establishment or new neighbor detected. For example, a router can send a Link Characteristics Request message to request metrics on the link from the modem. By replying with a Link Characteristics Response message the modem should send back at least the maximum and current data rate for up- and downlink and the maximum latency desired on the link. Additional information such as the link quality might also be added.

As can be seen, different approaches exist to determine link or path characteristics. Technology agnostic approaches, which evaluate TCP or perform active measurements, mostly aim at solely determining the available bandwidth of a connection in order to improve the TCP performance or to optimize capacity provisioning, routing and traffic engineering, respectively. Both are not per se suitable for converged terrestrial and satellite scenarios. Approaches that rely on link and physical layer evaluate technology specifics in order to determine the link characteristics. With DLEP an extensible protocol is now being standardized that allows for providing these link characteristics to higher layers.

Hence, the major challenge with respect to link characteristics identification is to design a method to continuously estimate the link characteristics beyond the available bandwidth of terrestrial and satellite links without significant overhead.

3) Path selection: Once the traffic's QoS requirements as well as the link characteristics are known, a proper path for the traffic needs to continuously be determined. While traditional IP routing is typically based on the destination IP and a simple metric, i.e. hop count or cost in general, the path selection method in converged satellite/terrestrial scenarios with parallel links needs to be more sophisticated. The traffic's QoS requirements need to be matched against the link characteristics in order to allow for a good end user QoE. Hence, not only the available capacity needs to be managed properly but also latency and jitter constraints. Moreover, the path selection needs to be on a more fine-grained basis compared to typical IP routing, since traffic exchanged between the same source and destination might be routed differently depending on the service application.

Particularly challenging in this context are the varying link conditions on satellite links. When the available capacity on a satellite link decreases, it needs to be checked if actions have to be taken in order to avoid a congestion. Furthermore, the different latency characteristics of satellite links also need to be considered. While in terrestrial networks low latency and low jitter values can be ensured by just avoiding congested links and, thus, preventing extensive queuing [69], in converged satellite and terrestrial networks the latency can be high even if sufficient capacity on the satellite link is available due to the high fixed latency, as explained earlier. The impact of the fixed latencies on the total transmission time depending on the amount of data to be sent is illustrated by an example depicted in Fig. 6. We calculate the theoretical time required to transmit data over different hypothetical links. Up to a certain threshold, which depends on the link speed, the total transmission time of a narrow-band hypothetical terrestrial link with negligible signal propagation time is faster than a high bandwidth satellite link. For an exemplary 2Mb/s terrestrial link this threshold is already reached when 60KBytes are in the queue. For a 10Mb/s terrestrial link around 500KBytes of queued data is required so that the satellite outperforms the terrestrial link. It should be noted that the MAC overhead is not taken into account in this figure and that 250ms and 20ms latencies are assumed for the satellite and the terrestrial link, respectively.

We would like to illustrate this issue further by the following example: A messenger pigeon carrying a 4GB USB drive over 63km takes around two hours [70]. This leads to a bandwidth of approximately 4Mbit/s, which might be even more than a current xDSL connection in some rural areas. Furthermore, while the QoS provided by this kind of network is most likely sufficient for large file transfer applications, it is absolutely unacceptable for any kind of application requiring interactivity.

As can be seen from this extreme example but also from Fig. 6, bandwidth and latency constraints need to be looked at together in scenarios with parallel satellite and terrestrial links. Even though the available capacity on a link might be lower, for a small amount of data a low latency link might be the better choice.

However, as explained earlier, not only do most TE concepts not consider such a scenario, but also common QoS mechanisms, such as Diffserv [71], neglect it. RFC3246 [72], which aims at providing a building block for low latency, jitter and loss services, indeed recognizes the fixed signal propagation time particularly on wide area links, yet the authors consider it as a "fixed property of the topology" and thus focus solely on minimizing the queuing latency. This is mainly due to the fact that the signal propagation times in terrestrial networks are, regardless of the actual technology, usually of the same order of magnitude and have less impact than the queuing delay. However, given the fact that for certain applications the overall latency has a similar high impact on the QoE as available capacity these approaches most likely lead to sub-optimal

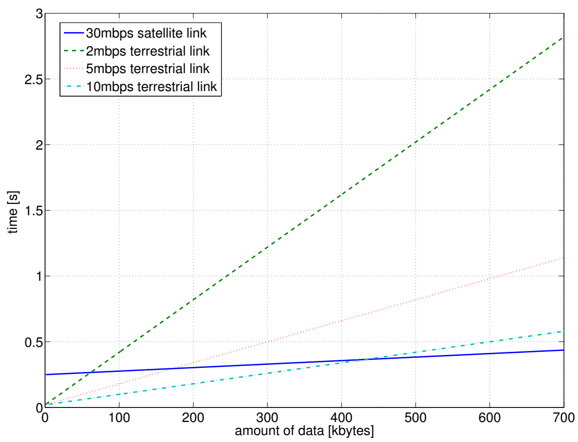


Fig. 6. Transmission time comparison of different terrestrial link speeds and 30Mbit/s satellite

results when integrated satellite and terrestrial networks with parallel links are used.

Given that, another major challenge to realize successfully converged satellite and terrestrial network scenarios with parallel links encompasses the definition of a proper path selection mechanism that is aware of and considers the special link characteristics of satellite links, including the high fixed latency. This mechanism has to perform regular TE tasks and beyond. That is, continuously managing the available capacity in the network, matching specific QoS requirements against the different link capabilities, so that optimal paths for all traffic flows (or traffic aggregates) are chosen.

C. Conclusion of technical challenges review

The challenges we mention in section IV-A are to a large extent already addressed in the literature, as we elaborate in section III. However, the challenges we identified with scenarios involving parallel terrestrial and satellite links as described in IV-B are only partially recognized but not yet addressed. In particular, an overall architecture is missing, which combines the different pieces efficiently. That is, it needs to be defined how existing mechanisms and techniques on various layers can usefully interact with each other. For example, a decision entity, which has to select to use either the terrestrial or the satellite link for a certain chunk of traffic, can perform a more informed decision if the capabilities and the condition of a link as well as the requirements of the traffic are known.

V. KEY BUILDING BLOCKS FOR CONVERGED SATELLITE AND TERRESTRIAL NETWORKS

The challenges presented in section IV-B can be, and to a certain extent already have been, addressed and solved independently for other use cases than integrating satellite networks with terrestrial ones. For instance, DPI traffic identification mechanisms are already productively in use for copyright content filtering or government surveillance [73]. These mechanisms are able to reliably identify applications being used in the network and can therefore help to derive their QoS demands. Similarly, methods are available that identify link characteristics, such as available bandwidth, in order to monitor the network or to enforce certain bandwidth guarantees.

However, even though most of the challenges we identified in the previous section seem to be solved individually in typical terrestrial networks, actual converged satellite and terrestrial networks, as described in section II, have not been realized. Besides economic constraints in the past, this is mainly due to two reasons: Firstly, not all mechanisms, which are designed for terrestrial networks, can cope with the specifics of satellite links and secondly, a holistic architecture is missing, which provides clear functional blocks and well-defined interfaces among them. That is, approaches providing a potential solution to the individual challenges are, on the one hand, required to cope with very heterogeneous links, and on the other hand need to be closely integrated with each other in order to operate jointly. This complex cross-layer information exchange becomes crucial in converged satellite and terrestrial networks due to the specific characteristics of the additional satellite links, which, if used not optimally, can dramatically decrease the user's QoE, even in lightly loaded networks. For example, if a specific chunk of traffic requires certain bandwidth and latency guarantees in order to provide the corresponding service with a high QoE, it is crucial that these requirements as well as the currently available capacity and latency values are known by the link selection process.

Consequently, functional key building blocks and their responsibilities as well as the information exchanged between them need to be defined, so that ultimately both terrestrial and satellite networks converge into a common architecture. This view is supported by the outcome of the literature review we performed in the previous section.

Hence, in this section, we give a high level overview of the key functional building blocks we identified as essential to realize a converged satellite and terrestrial network. The composition of these functional key building blocks is based on the previously presented challenges. Afterwards in section VI we map existing network architectures against these building blocks to evaluate their suitability to realize scenarios as described in section II.

Fig. 7 presents an overview of these key building blocks. Firstly, a Traffic Requirement Identification function is needed. Its responsibility is to determine, if applicable, the QoS requirements, such as maximum tolerable latency and jitter, priority, minimum required bandwidth, etc. per chunk of traffic that need to be fulfilled in order to achieve a high QoE. This information have to be provided to a TE component. For this purpose an interface, which we refer to as Ti, needs to be designed. This interface shall define how the traffic requirements can be described and how this information is exchanged. It should be noted that this functional block as well as the Ti interface also need to cope with dynamically varying requirements of a single application over time. For example, an FTP session has limited bandwidth requirements but certain latency requirements while directories are being browsed by a user. This significantly changes once an actual

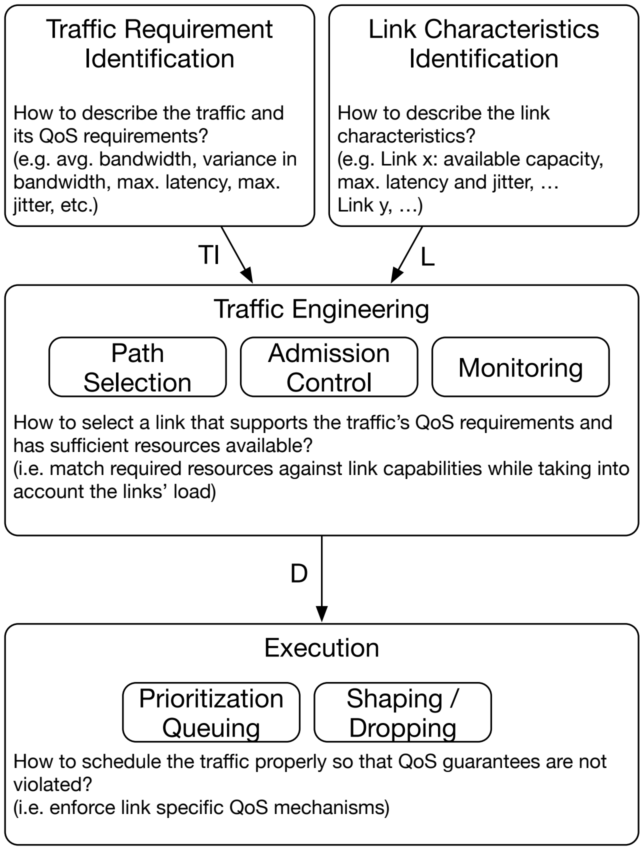


Fig. 7. Key functional building blocks

file transfer is started with high bandwidth demands but almost no latency requirements.

Secondly, a Link Characteristic Identification function is required, which determines and provides information on the different link characteristics of each available link to a TE component. This includes in particular the fixed latency values. Moreover, for capacity planing, additional information, such as available capacity, queue statuses, dynamic latencies or typical jitter, might also be determined to allow for a more informed path selection. It also needs to cope with dynamic link changes, which result from the satellite's ACM adaptions. Thus, we identify the need of a second interface, which we refer to as the L interface. The definition of the L interface shall include an accurate description of the link conditions as well as the means to convey them.

Thirdly, a TE functional block is necessary. This function considers all the available information and performs the task of TE and beyond. That is, it needs to verify that traffic can be admitted and plan the available capacities while considering the traffic's requirements. Latency and jitter demands need to be considered when path selection is performed and the traffic is distributed onto the different links, so that a high QoE is achieved. It should also be noted that the TE function is a continuous process, which requires monitoring of QoS guarantees, given that both the traffic's requirements as well as the link characteristics change over time.

In order to allow the TE component to execute its decision, we identify the necessity of a fourth functional building block, namely the Execution function, including another interface, which is referred to as the D interface. Its main purpose is to communicate the decision of the TE function to the lower network layers. The Execution block needs to perform traffic shaping and priority queuing based on the TE function's decision.

Furthermore, since the satellite has significant advantages with respect to broad- and multicast traffic, the key function building blocks should support the exploitation of this. That is, the number of users interested in a particular content as well as the link conditions to these users should be determined and communicated through the Ti and L interfaces. This way, a potential multicast group manager can decide to which users a particular content should be broadcasted via the satellite and to which ones it is better to send it terrestrially. For example, if a few users experience a bad satellite link quality due to rain, while others interested in the same content can be reached with a more efficient MCS, it might be worth to use the better MCS on the satellite transmission and send the content terrestrially to the users suffering from rain.

With respect to the OSI layers, the TE and the Traffic Requirement Identification functions need to operate on the higher layer, i.e. L3 and above, following a cross-layer approach. Instead of being located on a single OSI layer, information of different layers are essential to perform the aforementioned tasks properly. In contrast, the Execution function is located in layer 2, as it needs to control the actual access to the medium. The Link Characteristics Identification is either located in the lower layers, i.e. L1 and L2, if lower layer approaches are used, or on L4 and above if transport layer mechanisms or active measurements are used to determine the link characteristics.

To summarize, given the challenges we analyzed, we identified four key functional building blocks required to successfully integrate satellite and terrestrial networks, so that it is beneficial having parallel links. These are Traffic Requirement and Link Characteristics Identification, as well as TE and Execution function. Moreover, interfaces (i.e. Ti, L and D interfaces) are necessary to facilitate the interaction between the key functional building blocks, so that they can operate jointly.

VI. Applicability of Existing Network Architectures for Converged Satellite and Terrestrial Networks

We now evaluate different contemporary network architectures with respect to their feasibility to support converged FSS and terrestrial networks. Firstly, we look at a IP environments and known techniques. Secondly, we focus on current LTE approaches. Thirdly, we deal with DVB-S2/RCS2 based satellite architectures. Finally we analyze very recent alternative approaches, such as Software Defined Network (SDN) concepts or Wireless Back-haul (WiBACK).

A. IP QoS mechanisms

For many years now different mechanisms and technologies have been added to the IP ecosystem to deal with QoS, guaranteed services or to perform TE. Fig. 8 gives a non-exhaustive overview of this environment, which we discuss in the following.

1) Flow-based QoS approaches: Already more than one decade ago Quality-of-Service Open Shortest Path First (QoS-OSPF) [74], which is an extension to Open Shortest Path First (OSPF) [75], was defined to enable QoS in OSPF. It performs constraint-based routing, which in contrast to regular IP routing aims at finding a path in the network which fulfills certain bandwidth requirements, such as a minimum available bandwidth. This extension aims at enabling QoS routing in IP networks by distributing, along with the link states in the network, the available bandwidth for each link and by considering these in the routing decision. It should be noted that [74] does not consider any other QoS requirement apart from bandwidth. The Path calculation algorithm itself uses a modified Bellman-Ford algorithm which allows to determine paths of maximum available bandwidth for all hop counts by exploiting the fact that the Bellman-Ford algorithm progresses by increasing the hop count. Hence, among the paths which support the requested bandwidth, a path with the minimum number of hops can be selected without increasing the complexity. This limitation on the available bandwidth and hop count is caused by a general problem of constraint-based routing, namely its complexity. More complex metrics or a combination of metrics easily result in an NP-hard or even NP-complete problem. For example, selecting a path based on multiple independent QoS-constraints such as delay and cost has already proven to be a NP-complete problem, if the QoS metrics are real numbers or unbounded integers [76]–[78].

Furthermore, QoS-based routing is commonly proposed to be used together with a resource reservation protocol, such as RSVP [57], which blocks the required resource for a particular flow along the calculated path and therefore realize IntServ [59] in a non-overprovisioned network. As previously mentioned, applications on the end hosts systems (or intermediate router) can issue RSVP requests to gain specific qualities for a certain flow. These requests contain a flow description, e.g. source IP and port, and the desired QoS, and are routed from source to destination along the path calculated by the routing protocol. Each intermediate RSVP-enabled router on the path can then accept or reject the RSVP depending on its policies and available resources. Further details on RSVP can be found in [79]. It should be noted that IntServ can be used without QoS-OSPF but with any regular routing protocol. This, however, might lead to more denied requests if capacity constraints occur in the network, since the routing is performed unaware of the available resources on the router.

These approaches have some major drawbacks. As it is working flow-based and stateful, scalability can easily become an issue [80], regardless of the presence of satellite links in the network or not. Hence, neither IntServ nor QoS-OSPF are widely deployed in larger networks. Furthermore, the main idea behind these approaches is generally to have enough capacity available for critical traffic flows, since latency, jitter and loss usually increase if the available bandwidth is exceeded [69] or to prioritize real-time and other high-priority traffic by configuring the packet scheduler properly. Both are also important in converged satellite and terrestrial networks. However, those approaches do not consider the impact of the high fixed latency of satellite links, when choosing a path for certain traffic.

Even though [81] describes how network elements, i.e. routers, should behave to provide services with a guaranteed end-to-end delay and bandwidth, determining a path, which fits latency requirements is out of scope. Hence, satellite links are not treated properly, since a significant part of the latency is independent of the link's load. Moreover, while QoS-OSPF distributes link information within the network, it does not provide any means to gain information on the link characteristics, besides the link status, i.e. up or down.

Mapping these mechanisms against the key functional building block, as we described in section V, we can conclude that a limited Ti interface exists, i.e. RSVP. The Traffic Requirements Identification functional block can be seen as part of the application. However, the functional blocks of TE and Traffic Requirements Identification are not sufficient, since QoS-OSPF and RSVP only considers the bandwidth and other routing protocols are not QoS-aware. Furthermore, QoS-OSPF or other routing protocols determine the link state but not the actual characteristics. Thus, the L interface, which is inside the routing protocol, is also limited to the link state. Given that, the Link Characteristics Identification functional block is also not sufficient for providing a high QoE in converged satellite and terrestrial networks, since more than just the link state is needed. In this approach, the Execution functional block is implemented by an RSVP-enabled router and the corresponding D interface is implemented by RSVP.

2) Traffic-Aggregate based QoS: Given the scalability issues of flow-based approaches, current TE concepts usually operate on traffic aggregates. Already in the last century Diffserv [71], [82] has been standardized, which provides traffic classification and therefore allows for prioritization of IP traffic by using the IP Differentiated Services Code Point (DSCP) field. While various queuing strategies and optimizations exists (e.g. [83], [84]), Diffserv provides QoS only by prioritizing packets of higher priority traffic. It is only mentioned here for the sake of completeness, since pure prioritization is insufficient for converged satellite and terrestrial networks with parallel links due to the high fixed latency on satellite connections, as explained earlier.

Furthermore, current TE approaches often rely on Multi Protocol Label Switching (MPLS) Label-Switched Paths (LSPs) [85], [86] in order to become independent of the link layer technology. Instead of calculating a proper route for each flow, routes are calculated for an aggregate level on a larger-time scale. Hence, more complex path computation algorithms can be used compared to e.g. QoS-OSPF. Typically, RSVP-Traffic Engineering (RSVP-TE) [87] is used to establish these MPLS LSPs in the network and to block the proper resources [87], [88]. For different QoS classes different MPLS LSPs can be established, which introduces the challenge to map the traffic onto the LSPs.

Moreover, to further enhance the path computation, the IETF has created the so-called Path Computation Element (PCE) architecture [89]. Among other reasons, one moti-

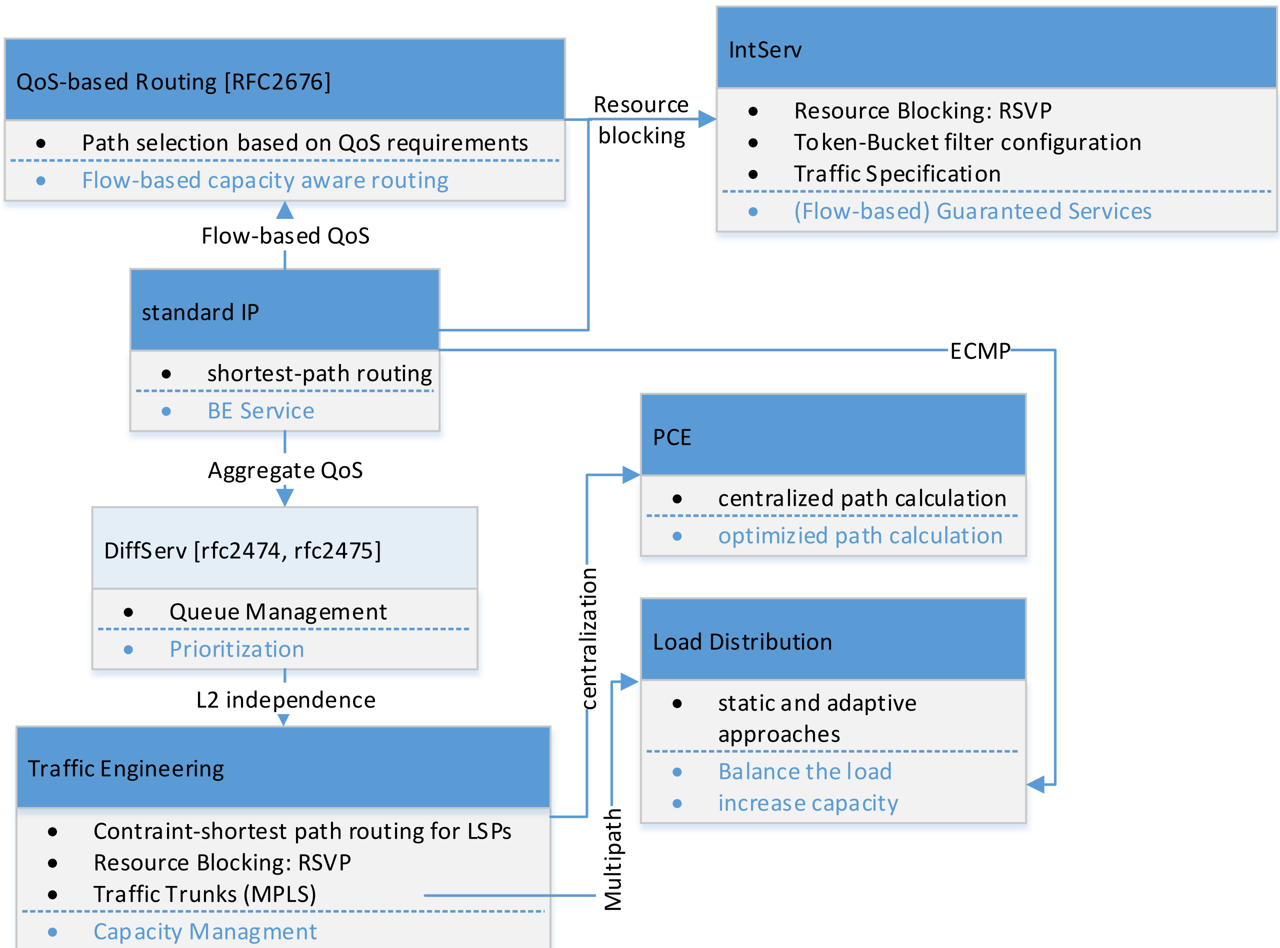


Fig. 8. IP TE environment

vation for PCE is to offload computational intensive path calculation, such as the path selection based on multiple independent constraints, which can easily overload a single network entity, to a more powerful machine. So-called Path Computation Clients (PCCs), which might be a router, can request a PCE to calculate a path. The PCE consists of a path computation module, a Traffic Engineering Database (TED) and a communication module. While the TED collects and tracks the current network status including the network topology and the link utilization, the path computation module calculates the requested paths based on the information from the TED and selected algorithms and policies. Unfortunately, it is not defined how the topology information is gathered, apart from that it should be learned from a routing protocol [90]. Moreover, with the PCE communication protocol and the communication module, a well-defined interface is available which enables PCCs to request path calculation between two arbitrary nodes in a standardized way. That is, a PCC, e.g. the first router, sends a PCReq message to a PCE optionally including among other things the required bandwidth or other metrics. The PCE replies with a PCRep message that includes the path, in case of a successful calculation, which, in turn, can be used by another signaling protocol, i.e. RSVP to set up a path in the network accordingly.

Furthermore, multiple LSPs might exist between any source and destination. In such a case the load must be distributed onto these paths. While there are specific MPLS approaches, such as [91] or [92], general algorithms exist as well. These can also be used with a regular IP routing protocol in case multiple paths of the same cost to a certain destination exist, which is referred to as Equal Cost Multipath (ECMP). The purpose of ECMP it to distribute traffic onto more than one path in order to balance the load and to avoid congestion, so that eventually a higher total throughput can be achieved [93]–[95]. A main differentiation between load distribution approaches is inter- and intra-connection parallelism. The traffic belonging to the same connection (or application) can be either sent over multiple available connections in parallel (intra-connection) or only over a single one and another connection might use a different parallel path (inter-connection). This is also often referred to as per-Packet distribution and per-Flow distribution, respectively. Per-packet approaches usually allow on the one hand for fine-grained and flexible path selection, since the link is selected for each packet independently, but on the other hand might cause re-ordering of packets. Re-ordering will occur particularly if the latency of the multiple links differ heavily, which for example impacts the performance of TCP [93] or introduces extra time needed to recover the packet order again [96]. In contrast, approaches providing inter-connection parallelism do not introduce re-ordering of packets of a single flow but in the case of very different flow characteristics might introduce over- or underloaded paths.

A further classification of load distribution into adaptive and non-adaptive models is given in [97]. While the non-adaptive models distribute the links statically and thus cannot react on dynamically changing conditions, adaptive models are able to react to variations in traffic- or network conditions. Typical examples for non-adaptive models are round-robin approaches or simple hash-based techniques.

Adaptive Load Balancing can be classified into Traffic-Condition-Based and Network-Condition-Based approaches. While Network-Condition-Based models can adapt to changing network conditions, such as delivery time or network utilization in terms of packets/s or bytes/s, Traffic-Condition-Based Models consider traffic characteristics, e.g. flow- or packet size, packet arrival time, etc., when performing load distribution. It should be noted that those classes are not mutually exclusive.

The Adaptive Flow-Level Load Control Scheme for Multipath Forwarding [98] is an example of a Traffic-Condition-Based approach. The authors make use of the fact that in IP networks one can generally differentiate between two kinds of flows, namely long-lived and short lived-flows. Short-lived flows, so-called transient flows, occur more frequently and have greater variation in packet arrival time than long-lived flows, while the long-lived flows carry the major part of the traffic load and, hence, are referred to as base flows. In order to detect a base flow the number of packets X per flow within a given time Y is measured. The idea of [98] is, if two links are available, to send all base flows over one path and all transient flows via the other one. By adapting X and Y the classification into base and transient flows can be dynamically changed and, thus, the assigned load can be controlled. The authors also propose a simple Load Control Algorithm, which adapts the X value based on the load ratio on the primary path. This is measured in packets sent via the primary path over the total number of packets.

An example for the Network-Condition-Based method is the Earliest Delivery Path First algorithm proposed in [99] and [100]. The authors' approach requires multiple connections between a mobile terminal and a counterpart, a so-called Network Proxy, in the network. By estimating the delivery time of each packet for each of the available paths, it schedules packets so that they are delivered as early as possible. The calculation takes into account the available bandwidth, the packet size and the queuing delay. This approach aims at increasing the usable bandwidth to the same performance a single link with the same aggregated bandwidth would have. However, [100] explicitly ignores the signal propagation delay, i.e. the fixed latency, since it is designed for terrestrial networks. Moreover, the Earliest Delivery Path First algorithm approach assumes that changes in the available capacity on each link and delay variations are only minor, which might not be an accurate assumption for satellite connections.

As can be seen, load distribution can be used to avoid congested links by balancing the load, even if the underlying networks are not homogeneous in terms of their capacity. However, traffic differentiation in terms of QoS requirements is not provided. Hence, load distribution needs to be done per traffic class. For further details please refer to [97], which provides a comprehensive study and comparison of different adaptive algorithms.

To conclude, IP-related TE approaches are highly optimized to perform TE in terrestrial networks as they aim at avoiding congested links, balancing the load, prioritizing certain traffic and managing the available capacity properly. However, the unique characteristics of satellite links are not taken into account - more precisely, the high fixed latency and the varying link conditions prohibit the integration of satellite links.

With respect to the key functional building blocks, we identified in section V, it can be seen that current IP QoS approaches operating on traffic aggregates cover them partially. The Traffic Requirement Identification functional building block is part of the application, which has to provide information on the traffic. A Ti interface is defined, which is realized either by RSVP messages and/or utilizing the IP DSCP field. Moreover, the Link Characteristics Identification functional building block is also included in the routing protocol. Hence, the link capabilities and conditions are not sufficiently determined. The TE function can be realized by PCE, which fosters sophisticated algorithms by allowing the offloading of complex calculations to powerful machines. Similar to IP flow-based QoS approaches, the Execution functional building block is implemented in RSVP-enabled MPLS devices, which perform traffic shaping and prioritize queuing.

B. LTE

LTE and Long Term Evolution Advanced (LTE-A) networks also share the challenge of providing a high service quality and even guaranteed services, such as Voice over LTE (VoLTE), to the MT or the end user, respectively. An LTE [101] network consists of an evolved Radio Access Network (eRAN) and an evolved Packet Core (EPC). While the first provides the actual radio access and the physical connection to the MTs via the eNodeBs, the latter is primarily responsible for maintaining IP connectivity. In order to allow for provisioning, so-called QoS bearers are established in the network, as specified in [102]. Each bearer is associated with a QoS Class of Identifier (QCI), which is a fixed mapping onto specific networks characteristics such as packet error rate or packet latency. For example, QCI index 1 maps on a maximum packet error loss rate of $10^{-}2$ packets and a packet delay budget of 100ms and, thus, is suitable for conversational voice traffic [103]. Individual flows are in turn mapped onto a specific bearer. Thus, all flows transported using the same bearer experience the same QoS treatment, such as shaping, prioritization, queue management or scheduling weights. Moreover, multiple bearers need to be established in oder to provide different QoS characteristics for different applications, e.g. for VoIP traffic a different bearer

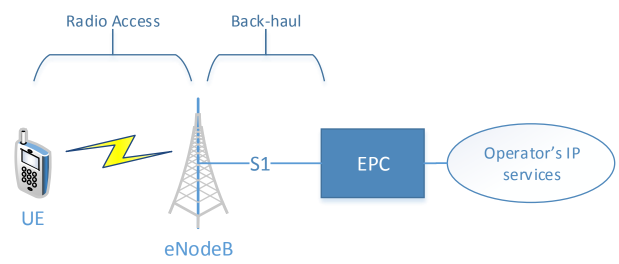


Fig. 9. Simplified LTE architecture overview

than for an File Transfer Protocol (FTP) download should be used.

The general QoS concepts of 3rd Generation Partnership Project (3GPP) networks are already explained in detail in many other publications, such as [104], [105]. It should be noted, however, since LTE is entirely an IP-based network, that it relies on its S1 interface [106], which connects the eNodeBs with the EPC, i.e. the back-haul network (see Fig. 9) on regular IP routing and the QoS mechanisms provided by the underlying network. In fact, there are no limitations on the used network technology [107]. Contrary to the radio access part where QoS enforcement mechanisms are included in the 3GPP standards, in the back-haul segment it is primarily the task of the underlying (heterogeneous) network to provide sufficient bandwidth or congestion control mechanisms, so that the required QoS levels can be guaranteed [108]. In fact, the back-haul is modeled as a point-to-point connection without resource contention [109]. In order to align the different QoS treatments, the bearer QCI can be mapped to the DSCP value of the IP packet, which in turn is used by the underlying network to provide proper QoS provisioning, or other IP-based QoS mechanisms should be used [110]. Currently the most common solution in operator networks is therefore to dimension the connection properly, so that a congestion does not occur or at least is highly unlikely. If the network cannot provide sufficient latency or capacity guarantees, [111] suggests several mitigation techniques, which aim at reducing the amount of traffic or changing priorities, e.g. by adapting used audio or video codecs. However, there are no means to allow the EPC e.g. to initiate a traffic re-routing or exploit multi-path connectivity. In fact, if satellite links are present the LTE architecture itself is not able to use them properly and to detect that only very relaxed latency guarantees can be given.

In this context, it is also logical that LTE networks have no mechanisms to monitor or determine the capabilities and properties of the back-haul network. That is, if multiple paths between the eNodeB and the EPC are available it cannot be decided by the EPC which one is used, it rather depends on the employed IP routing protocol. Moreover, LTE will not be aware of the high fixed latency of the satellite leading to delay intolerant bearers potentially being routed via the satellite.

With respect to traffic classification, 3GPP standards define a Traffic Detection Function (TDF). Its main goal is to detect the traffic of Over-the-top (OTT) applications such as NETFLIX, Facebook, etc., so that it can be charged and treated properly, i.e. assigned to a proper bearer [112]. TDF basically relies on DPI based traffic detection and provides a well-defined interface to other EPC modules, such as the Policy and Charging Rules Function (PCRF).

By mapping the LTE architecture against the previously identified key building blocks, it can be seen that LTE with its TDF provides a Ti interface and a framework for the Traffic Requirement Identification. However, neither the L nor the D interface exist, since the LTE architecture relies on IP and the underlying network technologies for QoS enforcement and routing in the back-haul network. Consequently, a TE functional block, which performs a path selection, is missing as well as an Execution function for the back-haul part of the network.

C. DVB-based architecture

Several approaches exist that enable QoS provisioning over DVB-S2/RCS satellite links. In the following we also analyze these with respect to their suitability to be used in scenarios with parallel satellite and terrestrial links. First and foremost, various standards and related associated guidelines exist for implementation that consider QoS in satellite networks, such as [113]–[115]. Their main motivation is to provision QoS by properly prioritizing traffic with more stringent latency requirements, to provide guaranteed bandwidth for certain services and to potentially perform admission control.

There is a strong focus on the interaction of layer 3 and layer 2, i.e. the satellite independent and satellite dependent layer. The QoS mechanisms of both layers need to work jointly together to enable QoS effectively. The main difficulties arise in particular on the return channel, when traffic is sent from the satellite terminal to the gateway. While on the forward channel it is mainly a scheduling problem, on the return link proper capacity needs to be requested due to the DAMA process. Moreover, as we describe in section IV-A2, different categories of requests are possible. Hence, it needs to be decided which kind of capacity is requested for which traffic. Unfortunately, how this mapping is done is outside of the scope of the standard.

The approach described in [113] relies on the Diffserv model [71] to classify and mark packets by using the IP DSCP field. The packets of the same class are assigned to so-called Behaivor Aggregates (BAs). All packets belonging to the same BA share the same network behavior. That is, they experience the same queuing and scheduling treatment. Depending on the BA, the DAMA process responsible for the medium access control requests and allocates capacity.

A concrete implementation for QoS provisioning over DVB-S2 is presented for example by [116]. By concatenating different round-robin schedulers and also considering the different MCS the authors implement fairness policies while giving QoS guarantees.

However, DVB-based architectures providing QoS aim at optimizing the QoS on satellite links. Unfortunately, they do not consider parallel terrestrial links, which could be exploited to increase the overall QoS and QoE. Hence, with respect to the key functional building blocks presented in section V, it can be seen that a Ti-interface exists, which is DSCP. Optionally a signaling proxy might be used, which, for instance snoops and interprets SIP messages. Given the scope of a DVB architecture, an actual TE function, which performs path selection and decides if traffic is routed terrestrially or via the satellite, is not part of this architecture. However, an Execution function, which is responsible for prioritization as well as bandwidth guarantees, is realized by proper scheduling and queuing mechanisms. Awareness of the link capabilities of the satellite link is implicitly available. [116] actually takes into account the different MCSs when scheduling and queuing packets. However, as terrestrial links are out of scope, only characteristics of the satellite link are available.

D. Alternative Architectures

In the following section, we evaluate emerging novel alternative approaches, with respect to their applicability for integrated satellite terrestrial networks. We firstly elaborate on architectures exploiting SDN concepts and secondly we focus on WiBACK, which is a network architecture specifically designed for heterogeneous wireless back-haul networks.

1) SDN approaches: It is generally believed that extended SDN approaches increase the flexibility of networks and enable novel concepts to address challenges in contemporary networks. With the advent of SDN [117] a paradigm change in networking architecture has started, shifting from monolithic network devices, which combine control, monitoring, management and data-forwarding functions in a single entity, towards a clear separation of control and data planes. That is, the decision making processes, such as routing of traffic, firewalling, spanning-tree protocols, etc., are clearly separated from the pure data forwarding methods. This allows for a more flexible management of the network, as the control functions can be run centralized. Further details on the differences between traditional networking and SDN are presented in [118].

SDN-enabled networks are mainly characterized by two things, first the decoupling of control- and data-plane, and second, programmability [119]. Fig. 10 shows the general SDN architecture [119]. At the lowest layer, the infrastructure layer, the actual data forwarding devices are located. Their main task is to perform any kind or packet processing based on the rules that the SDN-controller, which is located in the middle layer, provides. The functionality offered by the controller is often referred to as Network Operation System (NetOS). The most commonly used protocol between the SDN controller and the devices on the infrastructure layer is currently OpenFlow [120]. This interface is also often referred to as the Southbound interface. It is used to push rules to the infrastructure layer, to request monitoring information and statistics or to transmit packets, for which none of the rules apply to, back to the controller. Furthermore, the control layer provides an application programming interface (API), the so-called Northbound interface, to the application layer, which contains the so called network applications. An application might be as simple as a centralized Dynamic Host Configuration Protocol (DHCP) server or consist of more complex

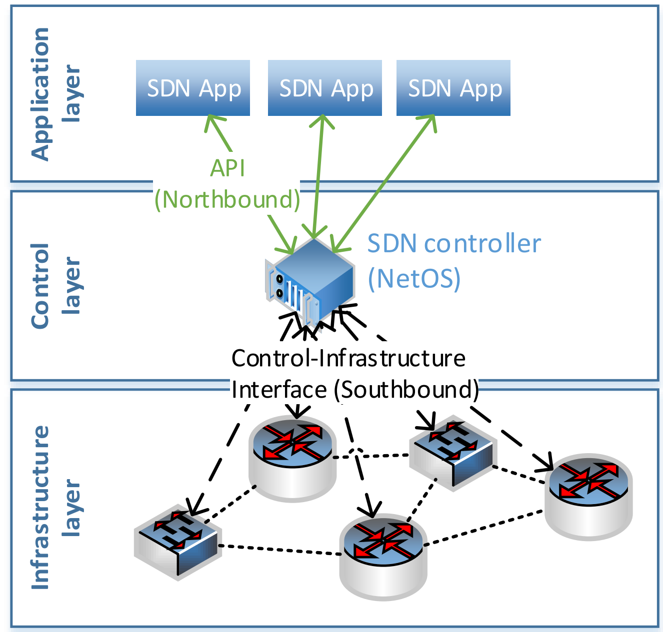


Fig. 10. SDN architecture

services like parental control for certain User Terminals (UTs) or seamless mobility. It should be noted that so far there is no standardized Northbound interface.

Several approaches exist to realize QoS provisioning by exploiting SDN concepts. An automatic QoS management mechanism for SDNs, which utilizes SDN concepts to automatically configure switches to provide QoS is proposed by [121]. The authors introduced a QoS control module acting as an SDN application. A context manager is responsible for gathering and aggregating information on the network, such as switch utilization or packet loss rate, as well as on the traffic flow. Based on these parameters, rule decision components check if QoS requirements can be satisfied and, if so, proper rules are implemented. These rules realize the enforcement of QoS by adjusting the scheduling queue on each affected switch as well as by properly classifying the packets. Unfortunately, [121] lacks a description on how information on the network status is gained from each device, since this is not part of the OpenFlow specification. Moreover, as the QoS management focuses only on queue management, high fixed latencies, as they occur on satellite links, are not considered by this approach.

Similarly, [122] exploits SDN concepts to deliver end-to-end QoS. The main idea of the authors is to differentiate between multimedia flows, which usually have stringent QoS requirements, and BE flows. While the first are routed using a dynamic QoS routing approach aiming at reducing the delay, the latter are routed using a typical shortest path first algorithm. SDN is used in particular to ease the route calculation by having the required information in a central point, namely the SDN controller, in [122]. However, similar to [121], the publication lacks mechanisms to determine latency, available bandwidth, etc. on each link.

Recently [123] acknowledges satellite and terrestrial network integration as a use case that can benefit from SDN approaches as they ease capacity aggregation, i.e. multi-link transmission, and load balancing. Moreover, the authors highlight data flow identification, link monitoring as well as dynamic forwarding rules generation and update as essential requirements for an SDN-based satellite/terrestrial integrated network solution. Unfortunately, the publication only outlines use-cases where SDN (and Network Function Visualization (NFV)) can be beneficial but concrete solutions are out of scope.

With respect to the key functional building blocks described in section V, it can be seen that the SDN architecture itself provides the D interface with the SDN Southbound interface, which allows for updating the flow tables of the network devices. Similarly an L interface can be realized. However, none of the approaches relying on SDN currently provide the actual building blocks, namely the Link Characteristics Identification, the Traffic Requirement Identification or a proper TE function, albeit the general SDN architecture eases the instantiation due to its centralized design. That is, all key functional building blocks can be implemented as SDN applications. It should be noted that the outcome of a recent European Space Agency (ESA) study [124] identified SDN as a key element for the successful integration of satellite and terrestrial networks, since interoperability and compatibility across the different segments is essential [125].

2) WiBACK: WiBACK [126] aims at providing a holistic cross-layer solution for wireless back-haul networks. It implements the concepts of SDN for data forwarding and also includes extensions beyond typical SDN to manage wireless interfaces. As depicted in Fig. 11, WiBACK adopts the general SDN architecture [117] and enhances it with wireless extensions on the Southbound interface. It also adds essential network applications performing capacity-aware routing, monitoring and link configuration, i.e. interface frequency and MCS assignment. Moreover, it defines a Northbound interface which allows for a flexible extension or creation of application.

The key parts of WiBACK are Spectrum Management and Capacity Management, both located in its control layer. The goal of Spectrum Management is to gain a global view on the physical network topology that identifies which interface (i.e. which (wireless) node) is physically able to communicate with which other interfaces of different devices. That is, it identifies which interfaces are of the same technology, can be tuned on the same frequency and are in communication range of each other. The algorithms of Spectrum Management are explained in detail in [127]. Spectrum Management selects the most optimal links out of the physically possible connections and creates a logical topology by configuring the wireless interfaces properly. Moreover, based on the selected MCS and the used Media Access Control (MAC) protocols on each link Spectrum Management calculates the available capacity on each link.

The Spectrum Management module is complemented by a Capacity Management module, which also operates in a centralized fashion, and performs the task of path calculation as well as resource, i.e. capacity, allocation. The required information, such as the network topology or the capacity on the individual links, is provided by Spectrum Management. Since Spectrum Management has created a logical topology, which consists only of the links that are already configured and whose capacity and typical link latency is determined, Capacity Management can operate without being aware of the actual technology, frequency or other physical layer parameter. Capacity Management itself is based on the concept of a centralized stateful PCE [89]. It relies on a shortest-path first algorithm which takes the available capacity on each link into account when paths of so-called pipes are calculated. Those pipes are MPLS label switched end-to-end paths which can be seen as traffic aggregates with an associated capacity, similar to LTE bearers. It should also be noted that WiBACK considers only the bandwidth as a QoS parameter. Other QoS parameters such as latency or loss are ignored when the path of a pipe is calculated. A more detailed discussion on management of the available capacity in WiBACK can be found in [128].

Furthermore, WiBACK has no means to identity the traffic QoS requirements. If the DSCP field of the IP header is properly set by the application, the flow is mapped onto a specific pipe. Currently WiBACK supports three types of pipes, which have a configurable set of QoS requirements associated with it. Those four pipe types are *Voice*, *Video*, *BE* and *Network Management*. The different pipe types are mapped in turn onto lower layer traffic classes, e.g. IEEE 802.11 Wireless Local Area Network (WLAN) access categories so that a prioritization can be realized. It should be noted that [129] has already shown how to integrate unidirectional overlay cells into the WiBACK architecture, which could be used to natively exploit the broadcast capabilities of the satellite network.

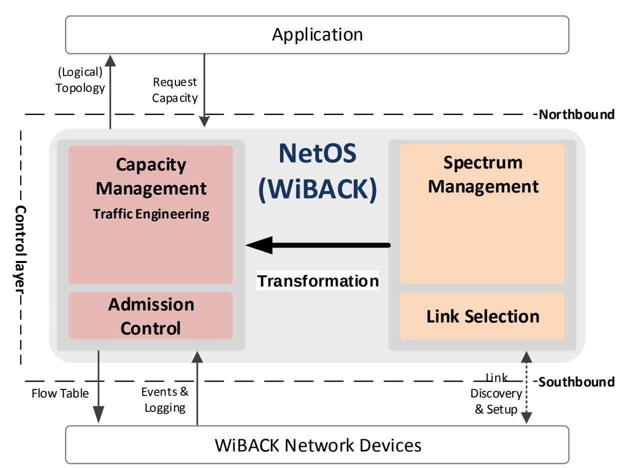


Fig. 11. WiBACK architecture

When comparing the key functional building blocks as presented in section V and the WiBACK architecture, it can be seen that WiBACK already provides a very sophisticated Link Characteristics block, which determines capacity, loss, latency and jitter of each link. Moreover, proprietary but well-defined L and D interfaces make this information available to the centralized Capacity Management and allows for decision enforcement, respectively. However, even though WiBACK's

Capacity Management module, which implements the TE functional building block, can access detailed link characteristics, it currently relies solely on the available capacity to calculate paths in the network. With respect to Traffic Requirement Identification, WiBACK relies on DSCP for the Ti interface in order to identify the Class of Service (COS) but does not implement the corresponding key building block.

VII. Comparison and gap analysis

In this section we compare the gaps that the aforementioned network architectures have with respect to convergence of satellite and terrestrial networks. This includes a comparison of the individual key features and unaddressed aspects, which are required to integrate parallel terrestrial and satellite links and, in particular, to deal with the high fixed latency on satellite connections.

As can be seen from the architecture review in section VI, which we summarize in Table II, elements of the key building blocks to successfully integrate satellite and terrestrial networks (see section V) are part of many architectures. However, not a single architecture sufficiently supports all identified key functional building blocks. Moreover, even though many aspects of the required key functional building blocks are already available in different networks architectures, it is undefined how the functions provided by different approaches interact. For example, it is yet undefined how a RSVP and PCE based IP approach could benefit from the link characteristics awareness of DVB-S2 satellite modem in order to realize packet delays QoS guarantees for LTE bearers. Due to missing interfaces, the awareness of the high fixed latency on the satellite link is not provided to the PCE by the DVB-S2 lower layers.

IP-based MPLS, PCE and RSVP are typical mechanisms used jointly in current operator networks to perform TE with good results. The generic PCE architecture enables offloading complex path calculation from network devices onto more powerful centralized computational nodes. This eases the exchange and optimization of the path calculation algorithms, which can easily access the topology information required from the TED. Hence, a proper algorithm taking the high fixed latency of satellite links adequately into account can be integrated without much effort. However, how to make the required information available at the TED is not clear. PCE can retrieve the topology itself from a routing protocol, which might include information on the available capacity, if provided by the routers, but usually neglects latency characteristics. Particularly given the more frequent changes in link characteristics in satellite networks, a tighter integration with the lower network layers is needed in any case to react faster to changes or failures.

The general LTE architecture is able to provision QoS and even provide guaranteed services by establishing bearers in the network that have certain QoS levels associated with it. While in the RAN segment the LTE radio enforces the guarantees for each bearer by scheduling packets properly, the back-haul segment relies upon IP and the underlying technology. That is, either sufficient capacity is always available, or the used network technology is able to prioritize packets properly, so that the QoS guarantees can be maintained. However, as IP is used, LTE shares the same drawbacks as IP-based approaches, which we explained above. First and foremost, high fixed latency links are not specifically taken into account and will contradict the QoS guarantees. Consequently, LTE does not provide any means to determine the link characteristics of the links in the back-haul segment, since it highly relies on the underlying IP routing. Hence, shifting delay tolerant traffic to the satellite network, i.e. certain non-guaranteed bearers, at the eNodeB, which is worthwhile to lower the load in the terrestrial back-haul networks, cannot be realized properly with the current LTE architecture.

Approaches based on DVB-S2/RCS2 solely focus on the satellite link. Consequently, a path selection distributing traffic to a parallel terrestrial connection is not considered and therefore not specified. Furthermore, awareness of the conditions of the satellite link is available and taken into account for packet prioritization or capacity request on the satellite link. With respect to a converged satellite and terrestrial network architecture supporting parallel links, those approaches can be logically located in the lower layers. That is, a TE function requires the information of the satellite link status and should also exploit the satellite's inherent scheduling and prioritization features. Unfortunately, there is no L interface defined that allows for exploiting this information by the higher layers. Such an interface is highly desirable so that IP-based approaches, e.g. PCE, can incorporate more accurate information in its decisions.

Emerging approaches to provide QoS by utilizing SDN concepts have the advantage of a centralized architecture and a controller that provides a global view on the network. Similarly to the PCE approach, this eases offloading of computational intensive path calculation algorithms. The key functional building blocks can be implemented as part of an SDN controller or as an network application. The implementation of the results of these algorithms in the network is also straightforward due to the design of the SDN architecture. Moreover, traffic identification might also benefit from the centralized SDN architecture, since computationally intensive classification algorithms can likewise be offloaded, if required. However, SDN does not perform any kind of Link Characteristics Identification function, even though the most-common SDN implementation OpenFlow can provide statistics on e.g. dropped packets or sent bytes, the latency and other link characteristics which are not being determined. In general it can be said, the SDN eases the control of IP packet flows in a network but SDN application that addresses the aforementioned drawbacks of IP-based approaches are still to be developed.

In contrast, the centralized WiBACK architecture focuses on accurately determining the capacity and other link characteristics by evaluating physical and link layer parameters, such as the MCS, and by passive monitoring. Furthermore, WiBACK establishes MPLS-based pipes in the network to block resources along a path. Similar to the LTE bearers, QoS parameters are associated with each pipe. In order to calculate the paths of the pipes WiBACK currently only considers the

Table II
OVERVIEW OF NETWORK ARCHITECTURES WITH RESPECT TO CONVERGED SATELLITE/TERRESTRIAL NETWORKS

Component	Relevant key features	Unaddressed aspects
QoS-OSPF [74] with RSVP [57]	• Calculation of shortest-path with sufficient bandwidth. • Resources (capacity) blocking along a calculated path. • Link up/down detection • Applications can signal required capacity. • Admission control.	• Just the link status (up/down) is detected but not more qualified ink characteristics, such as latency or jitter. • The main focus is to provide QoS guarantees by managing the available capacity. High fixed latencies on links negatively impacting the QoS are not considered. • Application cannot signal other QoS requirements than capacity.
MPLS LSP [85], [86] with RSVP-TE [87]	• Enhanced scalability due to operation on traffic aggregates. • Traffic can be prioritized. • Resources (capacity) blocking along a calculated path. • Applications can signal required capacity. • Admission control.	• Just the link status (up/down) is detected but not more qualified ink characteristics, such as latency or jitter. • The main focus is to provide QoS guarantees by managing the available capacity. High fixed latencies on links negatively impacting the QoS are not considered. • Application cannot signal other QoS requirements than capacity.
PCE [89]	• Offloading of path calculations to allows for more complex algorithms.	• A standardized way to gain link information required for the path calculations is missing.
ECMP approaches [97]–[100]	• Distributing traffic (packets or flows) onto multiple links.	• Distribution is not done based on QoS requirements.
LTE [101]	• Transmission of traffic with QoS guarantees. • Traffic identification mechanisms, i.e.TDF.	• Relies on IP-routing for path selection in the back-haul segment. Hence, LTE shares missing points with IP-approaches. • Congestion detection and other link estimation techniques are only defined for the RAN part. In the back-haul network LTE assumes a sufficiently provisioned network [107].
DVB-S2 [18], [113]	• Transportation of IP traffic via satellite connections. • Can provide detailed link status information on the satellite link.	• DVB-S2/RCS2 approaches assume all traffic being exchanged over the satellite link. Consequently, there is no path selection performed and the terrestrial link is not considered for traffic routing [113].
SDN [117]	• SDN provides a flexible architecture, which eases the implementation of the functional building blocks.	• On the southbound interface only very limited monitoring capabilities are available, e.g., OpenFlow provides counters for dropped packets or amount of received bytes. • Many required functionalities need to be implemented as SDN applications and are not yet available.
WiBACK [126]	• The Spectrum Management of WiBACK monitors lower layer link characteristics, so that ACM implications can be considered. • Block resources (capacity) along a calculated path.	• WiBACK calculates traffic paths only based on the available capacity. Other parameters are currently not considered [126].

available capacity on each link. Fortunately, WiBACK already determines the latency for each link, so that only the pipe calculation algorithm needs to be further enhanced to cope with the high fixed latency satellite links. With respect to the traffic QoS requirement identification, WiBACK does not provide the means to identify them but rather requires properly set DSCP header fields to map traffic onto the correct pipes based on the traffic's COS.

To conclude, existing architectures need to be enhanced or a novel architecture needs to be defined that supports converged satellite and terrestrial networks with parallel links. First and foremost, adaptations are needed that address the issues evolving from the high fixed latencies on satellite links. Such architectures are required to bring together the needed information to distribute traffic onto both networks, based on the QoS demands and networks capabilities. While IP-based approaches require modifications or enhancements in many protocols, e.g. as RSVP, routing protocols, PCE, exploiting and enhancing SDN concepts seem more promising and a cleaner approach to realize this due to their increased flexibility. This is also acknowledged by [125]. In particular, the WiBACK approach is an appealing option, since on the one hand, it adopts the SDN concepts, and on the other hand, it has a close interaction with the lower, technology-specific layers. We see this as a necessity in order to cope with the higher dynamics in satellite networks.

VIII. Key Performance Indicators to evaluate converged satellite and terrestrial networks

Having identified the key building blocks to instantiate converged satellite and terrestrial networks and matched those against current state-of-the-art network architectures, we now elaborate on the KPIs to compare and evaluate potential solutions. Meaningful and objective KPIs, which on the one hand are easily measurable, and on the other hand allow for assessing the performance of a specific approach, are essential to compare different approaches or to evaluate potential improvements. However, given the different key functional building blocks, which need to interact with each other, a wide variety of KPIs become important. Fig. VIII illustrates the different dimensions of KPIs relevant for converged satellite and terrestrial scenarios with parallel links including references to existing work that are discussed in the following paragraphs.

High user satisfaction is one of the most meaningful metrics, which is usually referred to as QoE perceived by the user. A common metric for QoE is the Mean Opinion Score (MOS), which describes the QoE on a scale from 1 to 5, where 1 referrers to the lowest quality and 5 to the highest [130]. Unfortunately the MOS is difficult to determine since subjective tests with test users and extensive test setups are required in order to gain reproducible and reliable values [131]. Determining the QoE by objective tests is highly difficult and application dependent. For example, in order to asses the video quality objectively, many specific metrics exist, starting from traditional SNR through to more sophisticated hybrids evaluating the video stream on various, application specific layers [132], [133]. However, these metrics are not applicable to determine the QoE of other applications, e.g. cloud applications.

Another major difficulty is that the QoE is influenced by several factors including but not limited to the type of device the user is using, the user profile, context, and also QoS parameters, such as throughput, latency, jitter or packet loss [134], [135]. The latter are mainly depending on the underlying network architecture and therefore are from a network perspective the ones with the highest impact factor on the QoE. However, the mapping between QoS and QoE is still an ongoing research topic [136]–[139]. In fact, [140] has shown that latency and jitter do not negatively impact the user's ability to assimilate the information yet it significantly impacts the QoE and, thus, it is essential to provide QoS at an appropriate level to achieve a high QoE [141].

Moreover, the different key functional buildings blocks also have dedicated KPIs depending on their individual task. For example, the quality of an algorithm identifying the traffic requirements can be evaluated based on its correctness and accuracy [142]. Similarly important is the accuracy when it comes to the identification of the link characteristics [143], [144]. In contrast, performance and scalability, e.g. execution time and space complexities, are highly relevant for the path selection process [145]. Furthermore, link utilization, reordering of packets or redistribution of flows display additional KPIs in multipath environments [97].

However, even though the individual performances of the different key functional building blocks alone clearly impact the overall system performance, it is insufficient to solely focus on the associated KPIs, since those do not consider the interaction between the key functional building blocks. For example, even if all key functional building blocks are operating ideally, the overall system performance might be negatively affected if timings between the different functions are not aligned, i.e. if the Path Selection Function is not aware of link characteristics or traffic requirements when it needs to make it decision, it cannot select a path optimally.

Given all that, focusing on the QoS parameter as KPIs to evaluate the overall system seems a promising option. These include throughput experienced by the user, the latency in terms of RTT or one-way latency, the packet delay variance (jitter) as well as the reliability in terms of packet- or Bit-error-rate [146]. Those KPIs can be either determined on a macro-level, considering the overall network, or on a micro-level, measuring individual flows. These different scopes are often referred to as system level performance or resource level performance, respectively [49]. For example, the throughput is either determined on a macro-level, considering the overall network, or on a micro-level, measuring individual flows. Both, system-level and resource-level, are equally important to determine the overall performance of a system.

To conclude, the goal of an evaluation process is the assessment of the quality of an overall architecture including the performance of its key functional building blocks and the overall system behavior. Given that the main goal of integrating satellite and terrestrial networks is to provide a joint network, which is able to cope with the bandwidth demands of future networks while ideally achieve the same QoE as in high-bandwidth terrestrial connections, QoE is a promising KPI candidate. Unfortunately it is difficult to measure, which makes it essential to use other KPIs. However, neither single QoS KPIs, such as throughput or latency, nor the independent evaluation of the algorithms forming the key building blocks is sufficient. Moreover, different trade-offs might have to be made during the instantiation of a real system. Hence, multiple KPIs needs to be considered, weighted and evaluated in different situations to holistically determine the performance a potential solution.

IX. Future research directions

After analyzing the current state of the art with respect to convergence of satellite and terrestrial networks, we identified several open research questions, which should be further discussed to foster the successful integration of satellite and terrestrial networks, as we envisioned in section II.

It should be noted that we focus on the open issues affecting the instantiation of converged satellite and terrestrial network scenarios. Future research directions concerning the individual key building blocks have already been discussed in other publications, such as [54], [142], [147], and are therefore not repeated here.

Moreover, we distinguish between architectural and evaluation issues. While the former raises open questions with respect to the specification of the architecture, the latter deals with future research aspects regarding evaluation.

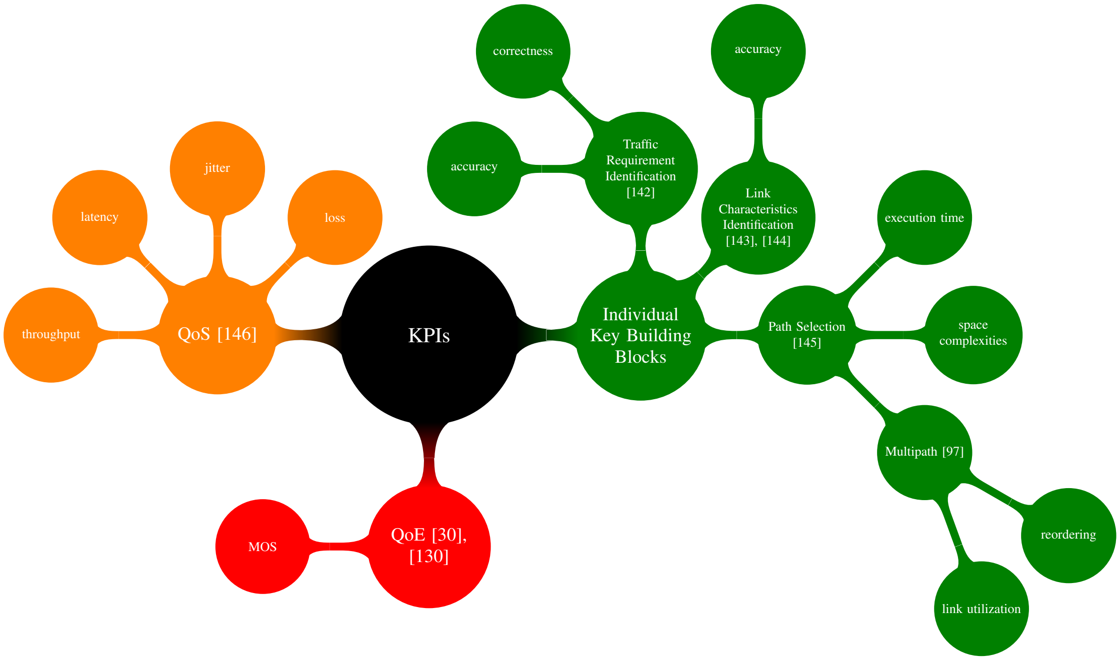


Fig. 12. Dimensions of KPIs

A. Architectural

- How much dynamism is required?
  A relatively simple instantiation of converged satellite and terrestrial networks with parallel connections might rely on the IP DSCP field and a path selection algorithm, which routes background traffic and other classes without strict latency demands via the satellite and network control and real-time traffic terrestrially. Such a highly static approach does not introduce much computational overhead, assuming the IP DSCP field is properly set by the application or can be easily determined by other means. Furthermore, the requirements on the Link Characteristics Identification as well as TE and Execution function are also limited. The complexity increases once the path selection decides dynamically based on the current load of the different links, which classes are routed via the satellite or via the terrestrial links. For instance, if the utilization of terrestrial links is low, more traffic (classes) might be forwarded terrestrially. However, this dynamism comes with the cost of increased complexity. Hence, what level of dynamism is beneficial for the overall system performance needs to be investigated. This includes identifying the most relevant parameters for the path selection.
- What level of accuracy and which parameters are required to be determined by the Link Capability Identification?
  A minimalist approach might solely differentiate between satellite links and terrestrial ones without further characterizing them, while more complex techniques might determine more parameters of the link, such as available capacity or jitter. Moreover, estimation methods operating with lower layer information usually detect changes on a link faster but are also more complex, since they require knowledge about the underlying technology, whereas e.g. transport layer based methods are technology agnostic and detect changes by the reaction of the transport layer protocol. Hence, one needs to evaluate in which situations the QoE significantly increases if more informed or faster decisions can be made, so that the additional complexity has a real benefit.
- How many details with respect to traffic requirements identification are useful?
  A relatively simplistic approach might classify traffic into two categories, namely non-interactive traffic that can tolerate the high fixed latency of satellite traffic and interactive traffic, which needs to be delivered as quickly as possible. On the other hand, a more complex Traffic Requirements Identification algorithm might be able to determine more precisely the actual demands of a flow in terms of necessary bandwidth, maximum loss, latency and jitter. While the first approach also leads to a simplified path selection algorithm, since there is only a single parameter available based on which the path is being selected, the latter allows for a more complex selection process, including that the requirements of an application change over time. Particularly in overload situations this might allow for a more informed decision. Given that, what should be identified is, which parameters bring a clear benefit and which disproportionately increase the overhead.

B. Evaluation

- What is a representative baseline instantiation?
  A major challenge is that to the best of the authors' knowledge there are no deployments of converged satellite and terrestrial networks complementing each other so that multiple paths between two nodes in the network exist, as we describe in section II. Thus, there

is no implementation which can be used as a baseline, against which developed approaches or improvements can be compared to. Thus, any converged satellite/terrestrial solution needs to be thoroughly compared with pure terrestrial or pure satellite deployments. That is, if a terrestrial network is assumed, what is the benefit that an additional satellite network providing a parallel connection could bring? Unfortunately, this depends highly on the concrete deployment scenario, e.g. in a rural region with only low capacity links, the gain of additional high capacity satellite links might be much higher compared to that of well-connected urban areas, where optical fiber are deployed at a large-scale. Similarly, the predominant applications being used are also relevant to assess the benefit. If the majority of traffic expects low response times the additional satellite links will only bring limited benefit to the overall network performance.
Hence, representative scenarios need to be defined to compare the performance of an integrated satellite/terrestrial network with a typical terrestrial-only deployment. This includes, among others, defining the number of nodes in the network, typical capacities, amount of users as well as used applications, so that representative use cases are tested.

- What are relevant KPIs?
  As explained in section VIII, selecting meaningful and easily measurable KPIs is not a trivial task, given that a huge variety of different KPIs exist and the QoE is difficult to measure. When novel network architectures or technologies, such as LTE, LTE-A or Very-high-bit-rate digital subscriber line (VDSL), have been introduced, often the throughput per user or the total system throughput is used to quantify the improvements of the respective architecture. As has already been mentioned multiple times, due to the unique link characteristics of satellite links, focusing only on the throughput will display a biased view, since the high fixed latency of satellite systems is ignored. However, even if throughput and latency are both measured, the results need to carefully be evaluated, since their relevance depends highly on the traffic that is transported over the network. Thus, a set of KPIs needs to be determined and, if necessary, properly weighted in order to allow for comprehensive evaluation results.

X. Conclusions

We have analyzed the current state of the art with respect to satellite and terrestrial network convergence. In this work we focus on scenarios in which emerging and more efficient satellite networks are used jointly with terrestrial networks, providing alternative parallel links and capacity virtually everywhere.

After defining concrete convergence scenarios and analyzing the related work, we elaborate on the technical challenges of satellite and terrestrial convergence. Since the characteristics of satellite links are highly different, particularly in terms of signal propagation time and therefore the link latency, several technical challenges need to be solved so that satellite links can successfully complement terrestrial connections. First and foremost one must to ensure that only traffic is routed via the satellite link, which can cope with its link characteristics. Thus, we identified several key functional building blocks required to instantiate an architecture for a successful convergence of satellite and terrestrial networks. The first block is a Traffic Requirement Identification function responsible for determining the QoS requirements of traffic. Second, a Link Characteristics Identification function, which has the task of providing information on the link capabilities, such as available capacity, loss, jitter or latency is also needed. Third, is a TE function, which maps the traffic requirements against the link capabilities. The final block is an Execution function that performs traffic shaping and prioritization. We argue that these four blocks are crucial to realize a successful convergence. Hence, we analyzed current state-of-the-art network architectures with respect to these building blocks. Unfortunately, none of the considered architectures implement all key functional building blocks sufficiently, so that the convergence of satellite and terrestrial networks has not happened and further work is required.

Furthermore, we elaborate on meaningful KPIs that can help to evaluate different concrete approaches. We identified the QoE as the most meaningful metric to asses the quality of an overall system. However, since it requires significant effort to measure the QoE, alternative KPIs need to be selected to allow for a thorough evaluation. Here QoS related parameters, such as latency, throughput, loss and jitter, seem most promising for the given use cases.

Finally, we identified several concrete open research issues that are required to be solved in order to realize converged satellite and terrestrial networks with parallel connections.

References

[1] Cisco Cooperation, "The Zettabyte Era-Trends and Analysis," Whitepaper, 2013. [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/VNI_Hyperconnectivity_WP.pdf

[2] Network2020 ETP, "5G: Challenges, Research Priorities, and Recommendations," Whitepaper, September 2014. [Online]. Available: http://networld2020.eu/wp-content/uploads/2015/01/Joint-Whitepaper-V12-clean-after-consultation.pdf

[3] European Commission - Directorate-General for Communication Networks, Content and Technology (CONNECT), "Digital agenda for europe scoreboard 2012," June 2012. [Online]. Available: http://ec.europa.eu/digital-agenda/sites/digital-agenda/files/KKAH12001ENN-PDFWEB_1.pdf

[4] P. Briggs, R. Chundury, and J. Olsson, "Carrier ethernet for mobile backhaul," *IEEE Commun. Mag.*, vol. 48, no. 10, pp. 94–100, October 2010.

[5] D. Henkel *et al.*, "Connecting the unconnected - economic constraints and technical requirements towards a back-haul network for rural areas," in *Proc. Global Commun. Conf. Workshops*, Dec 2011, pp. 1039–1044.

[6] P. Rost *et al.*, "Cloud technologies for flexible 5g radio access networks," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 68–76, May 2014.

[7] C. Niephaus, M. Kretschmer, and G. Ghinea, "Towards qos-aware load distribution in heterogeneous networks," in *Proc. IEEE 11th Malaysia Int. Conf. on Commun.*, Nov 2013, pp. 151–156.

[8] B. Evans *et al.*, "Integration of satellite and terrestrial systems in future multimedia communications," *IEEE Wireless Commun.*, vol. 12, no. 5, pp. 72–80, Oct 2005.

[9] ——, "Concepts and technologies for a terabit/s satellite," in *Proc. 3rd Int. Conf. on Advances in Satellite and Space Communications*, Budapest, Hungary, 2011.
[10] A. Botta and A. Pescape, "New generation satellite broadband internet services: Should adsl and 3g worry?" in *IEEE 32nd Int. Conf. Comput. Commun. Workshops*, April 2013, pp. 399–404.
[11] Y. Chen, T. Farley, and N. Ye, "Qos requirements of network applications on the internet," *Inf. Knowl. Syst. Manag.*, vol. 4, no. 1, pp. 55–76, 2004.
[12] *Internet protocol aspects – Quality of service and network*, ITU-T Recommendation Y.1540, 2011.
[13] *Network Performance Objectives for IP-Based Services*, ITU-T Recommendation Y.1541, 2003.
[14] "Ericsson Mobility Report," Whitepaper, November 2014. [Online]. Available: http://www.ericsson.com/res/docs/2014/ericsson-mobility-report-november-2014.pdf
[15] E. Bastug, M. Bennis, and M. Debbah, "Living on the edge: The role of proactive caching in 5g wireless networks," *IEEE Commun. Mag.*, vol. 52, no. 8, pp. 82–89, Aug 2014.
[16] A. Lykourgiotis *et al.*, "Hybrid broadcast and broadband networks convergence for immersive tv applications," *IEEE Wireless Commun.*, vol. 21, no. 3, pp. 62–69, June 2014.
[17] Q. Zhang, J. Chen, and H. Zhu, "Network convergence: theory, architectures, and applications," *IEEE Wireless Commun.*, vol. 21, no. 6, pp. 48–53, December 2014.
[18] *Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications (DVB-S2)*, ETSI EN 302 307, Rev. V1.3.1, March 2013.
[19] *Interaction channel for Satellite Distribution Systems*, ETSI EN 301 790, Rev. V1.5.1, May 2009.
[20] M. Breiling *et al.*, "Lte backhauling over meo-satellites," in *Proc. 7th Adv. Satellite Multimedia Syst Conf. and 13th Signal Process. for Space Commun. Workshop*, Sept 2014, pp. 174–181.
[21] S. Kota and M. Marchese, "Quality of service for satellite ip networks: a survey," *Int. J. of Satellite Commun. and Netw.*, vol. 21, no. 4-5, pp. 303–349, 2003.
[22] F. Zampognaro, "Satellite and terrestrial network integration," Ph.D. dissertation, University of Rome Tor Vergata, 2009.
[23] T. Gayraud and P. Berthou, "A qos architecture for dvb-rcs next generation satellite networks." *EURASIP J. Wireless Commun. and Netw.*, vol. 2007, 2007.
[24] L. Audah, Z. Sun, and H. Cruickshank, "Qos evaluation of multiservice applications over integrated satellite-terrestrial networks," in *Proc. 3rd Int. Congr. on Ultra Modern Teleccommun. and Control Syst. and Workshops*, Oct 2011, pp. 1–8.
[25] I. Akyildiz and S.-H. Jeong, "Satellite atm networks: a survey," *IEEE Commun. Mag*, vol. 35, no. 7, pp. 30–43, Jul 1997.
[26] G. Peters *et al.*, "The bats project," in *Proc. Future Netw. and Mobile Summit*, 2013, pp. 1–8.
[27] S. Watts and O. Aliu, "5g resilient backhaul using integrated satellite networks," in *Proc. 7th Adv. Satellite Multimedia Syst Conf. and 13th Signal Process. for Space Commun. Workshop*, Sept 2014, pp. 114–119.
[28] "Cost benefit analysis," Deliverable, EU BATS Project, April 2014. [Online]. Available: http://batsproject.eu/images/PublicDeliverables/BATS_D5.2_AVA_v1_F.pdf
[29] "5g radio network architecture," Whitepaper, Radio Access and Spectrum FP7 Future Networks Cluster, March 2014.
[30] G. Fairhurst *et al.*, "Delivery of triple-play services over broadband satellite networks," *IET Commun.*, vol. 4, no. 13, pp. 1544–1555, September 2010.
[31] H. Skinnemoen *et al.*, "Voip over dvb-rcs with qos and bandwidth on demand," *IEEE Wireless Commun.*, vol. 12, no. 5, pp. 46–53, Oct 2005.
[32] L. Siyang *et al.*, "Lte-satellite: Chinese proposal for satellite component of imt-advanced system," *China Communications*, vol. 10, no. 10, pp. 47–64, Oct 2013.
[33] M. Amadeo *et al.*, "A satellite-lte network with delay-tolerant capabilities: Design and performance evaluation," in *Proc. IEEE Veh. Technol. Conf.*, Sept 2011, pp. 1–5.
[34] V. Cerf *et al.*, "Delay-Tolerant Networking Architecture," RFC 4838 (Informational), Internet Engineering Task Force, Apr. 2007. [Online]. Available: http://www.ietf.org/rfc/rfc4838.txt
[35] C. Caini, R. Firrincieli, and D. Lacamera, "Pepsal: A performance enhancing proxy for tcp satellite connections [internetworking and resource management in satellite systems series]," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 22, no. 8, pp. B–9–B–16, Aug 2007.
[36] P. Arapoglou *et al.*, "Mimo over satellite: A review," *IEEE Commun. Surveys Tuts.*, vol. 13, no. 1, pp. 27–51, First 2011.
[37] I. Akyildiz, J. Xie, and S. Mohanty, "A survey of mobility management in next-generation all-ip-based wireless systems," *IEEE Wireless Commun.*, vol. 11, no. 4, pp. 16–28, Aug 2004.
[38] F. Arnal *et al.*, "Handover management for hybrid satellite/terrestrial networks," in *Personal Satellite Services*, ser. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, P. Pillai, R. Shorey, and E. Ferro, Eds. Springer Berlin Heidelberg, 2013, vol. 52, pp. 95–103.
[39] *IEEE Standard for Local and metropolitan area networks - Media Independent Handover Services*, IEEE Std. 802.21-2008, 2009.
[40] J.-M. Park, D.-S. Oh, and D.-C. Park, "Coexistence of mobile-satellite service system with mobile service system in shared frequency bands," *IEEE Trans. Consum. Electron.*, vol. 55, no. 3, pp. 1051–1055, August 2009.
[41] V. Deslandes, J. Tronc, and A.-L. Beylot, "Analysis of interference issues in integrated satellite and terrestrial mobile systems," in *Proc. 5th Adv. Satellite Multimedia Syst Conf. and 11th Signal Process. for Space Commun. Workshop*, Sept 2010, pp. 256–261.
[42] T. Taleb, Y. Hadjadj-Aoul, and T. Ahmed, "Challenges, opportunities, and solutions for converged satellite and terrestrial networks," *IEEE Wireless Commun.*, vol. 18, no. 1, pp. 46–52, February 2011.
[43] S. Kota, G. Giambene, and S. Kim, "Satellite component of ngn: Integrated and hybrid networks," *Int. J. of Satellite Commun. and Netw.*, vol. 29, no. 3, pp. 191–208, 2011.
[44] M. Allman, V. Paxson, and E. Blanton, "TCP Congestion Control," RFC 5681 (Draft Standard), Internet Engineering Task Force, Sep. 2009. [Online]. Available: http://www.ietf.org/rfc/rfc5681.txt
[45] D. Papadimitriou *et al.*, "Open Research Issues in Internet Congestion Control," RFC 6077 (Informational), Internet Engineering Task Force, Feb. 2011. [Online]. Available: http://www.ietf.org/rfc/rfc6077.txt
[46] J. Border *et al.*, "Performance Enhancing Proxies Intended to Mitigate Link-Related Degradations," RFC 3135 (Informational), Internet Engineering Task Force, Jun. 2001. [Online]. Available: http://www.ietf.org/rfc/rfc3135.txt
[47] *DVB-RCS2 Lower Layer Satellite Specification*, ETSI EN 301 545-2, Rev. V1.2.1, April 2014.
[48] G. Gardikis and A. Kourtis, "Using dvb-s2 adaptive coding and modulation for the provision of satellite triple play services," *IEEE Commun. Mag.*, vol. 46, no. 12, pp. 128–135, December 2008.
[49] D. Awduche *et al.*, "Overview and Principles of Internet Traffic Engineering," RFC 3272 (Informational), Internet Engineering Task Force, May 2002, updated by RFC 5462. [Online]. Available: http://www.ietf.org/rfc/rfc3272.txt
[50] Y. Lee and B. Mukherjee, "Traffic engineering in next-generation optical networks," *IEEE Commun. Surveys Tuts.*, vol. 6, no. 3, pp. 16–33, Third 2004.
[51] N. Wang *et al.*, "An overview of routing optimization for internet traffic engineering," *IEEE Commun. Surveys Tuts.*, vol. 10, no. 1, pp. 36–56, First 2008.
[52] P. Fan *et al.*, "Application Enabled Collaborative Networking: Problem Statement," draft-conet-aeon-problem-statement-01, July 2014. [Online]. Available: https://tools.ietf.org/html/draft-conet-aeon-problem-statement-01
[53] L. Bernaille *et al.*, "Traffic classification on the fly," *SIGCOMM Comput. Commun. Rev.*, vol. 36, no. 2, pp. 23–26, Apr. 2006.
[54] T. Nguyen and G. Armitage, "A survey of techniques for internet traffic classification using machine learning," *IEEE Commun. Surveys Tuts.*, vol. 10, no. 4, pp. 56 –76, quarter 2008.
[55] A. Dainotti, A. Pescape, and K. Claffy, "Issues and future directions in traffic classification," *IEEE Netw.*, vol. 26, no. 1, pp. 35 –40, january-february 2012.
[56] S. Jiang *et al.*, "Application enabled collaborative networking gap analysis," draft-conet-aeon-gap-analysis-01, October 2014. [Online]. Available: https://tools.ietf.org/id/draft-conet-aeon-gap-analysis-01.txt
[57] R. Braden *et al.*, "Resource ReSerVation Protocol (RSVP) – Version 1 Functional Specification," RFC 2205 (Proposed Standard), Internet Engineering Task Force, Sep. 1997, updated by RFCs 2750, 3936, 4495, 5946, 6437, 6780. [Online]. Available: http://www.ietf.org/rfc/rfc2205.txt
[58] J. Manner *et al.*, "Using and Extending the NSIS Protocol Family," RFC 5978 (Informational), Internet Engineering Task Force, Oct. 2010. [Online]. Available: http://www.ietf.org/rfc/rfc5978.txt
[59] J. Wroclawski, "The Use of RSVP with IETF Integrated Services," RFC 2210 (Proposed Standard), Internet Engineering Task Force, Sep. 1997. [Online]. Available: http://www.ietf.org/rfc/rfc2210.txt

[60] F. L. Faucheur *et al.*, "Resource Reservation Protocol (RSVP) Proxy Approaches," RFC 5945 (Informational), Internet Engineering Task Force, Oct. 2010. [Online]. Available: http://www.ietf.org/rfc/rfc5945.txt

[61] ——, "Resource Reservation Protocol (RSVP) Extensions for Path-Triggered RSVP Receiver Proxy," RFC 5946 (Proposed Standard), Internet Engineering Task Force, Oct. 2010. [Online]. Available: http://www.ietf.org/rfc/rfc5946.txt

[62] J. Rosenberg *et al.*, "SIP: Session Initiation Protocol," RFC 3261 (Proposed Standard), Internet Engineering Task Force, Jun. 2002, updated by RFCs 3265, 3853, 4320, 4916, 5393, 5621, 5626, 5630, 5922, 5954, 6026, 6141, 6665, 6878. [Online]. Available: http://www.ietf.org/rfc/rfc3261.txt

[63] M. Handley, V. Jacobson, and C. Perkins, "SDP: Session Description Protocol," RFC 4566 (Proposed Standard), Internet Engineering Task Force, Jul. 2006. [Online]. Available: http://www.ietf.org/rfc/rfc4566.txt

[64] W. George *et al.*, "Application Enabled Collaborative Networking Use Cases," draft-conet-aeon-use-cases-01, July 2014. [Online]. Available: https://tools.ietf.org/html/draft-conet-aeon-use-cases-01

[65] K. Xu, Y. Tian, and N. Ansari, "Tcp-jersey for wireless ip communications," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 4, pp. 747–756, May 2004.

[66] M. Jain and C. Dovrolis, "End-to-end available bandwidth: measurement methodology, dynamics, and relation with tcp throughput," *IEEE/ACM Trans. Netw.*, vol. 11, no. 4, pp. 537–549, Aug 2003.

[67] W. Ivancic *et al.*, "Modem link-property advertisements," in *Proc. IEEE Aerospace Conference*, 2012, pp. 1–7.

[68] S. Ratliff *et al.*, "Dynamic Link Exchange Protocol (DLEP)," draft-ietf-manet-dlep-04, March 2013. [Online]. Available: http://tools.ietf.org/html/draft-ietf-manet-dlep-04

[69] X. Xiao and L. Ni, "Internet qos: a big picture," *IEEE Netw.*, vol. 13, no. 2, pp. 8–18, Mar 1999.

[70] The pigeon versus the computer: A surprising win for all. [Online]. Available: http://www.pigeonrace2009.co.za/

[71] S. Blake *et al.*, "An Architecture for Differentiated Services," RFC 2475 (Informational), Internet Engineering Task Force, Dec. 1998, updated by RFC 3260. [Online]. Available: http://www.ietf.org/rfc/rfc2475.txt

[72] B. Davie *et al.*, "An Expedited Forwarding PHB (Per-Hop Behavior)," RFC 3246 (Proposed Standard), Internet Engineering Task Force, Mar. 2002. [Online]. Available: http://www.ietf.org/rfc/rfc3246.txt

[73] R. Bendrath, "Global technology trends and national regulation: Explaining Variation in the Governance of Deep Packet Inspection," Paper prepared for the International Studies Annual Convention New York City, February 2009. [Online]. Available: http://userpage.fu-berlin.de/~bendrath/Paper_Ralf-Bendrath_DPI_v1-5.pdf

[74] G. Apostolopoulos *et al.*, "QoS Routing Mechanisms and OSPF Extensions," RFC 2676 (Experimental), Internet Engineering Task Force, Aug. 1999. [Online]. Available: http://www.ietf.org/rfc/rfc2676.txt

[75] J. Moy, "OSPF Version 2," RFC 2328 (INTERNET STANDARD), Internet Engineering Task Force, Apr. 1998, updated by RFCs 5709, 6549, 6845, 6860. [Online]. Available: http://www.ietf.org/rfc/rfc2328.txt

[76] S. Chen and K. Nahrsted, "An overview of quality of service routing for next-generation high-speed networks: problems and solutions," *IEEE Netw.*, vol. 12, no. 6, pp. 64–79, 1998.

[77] F. Kuipers *et al.*, "An overview of constraint-based path selection algorithms for qos routing," *IEEE Commun. Mag*, vol. 40, no. 12, pp. 50–55, 2002.

[78] X. Yuan, "Heuristic algorithms for multiconstrained quality-of-service routing," *IEEE/ACM Trans. Netw.*, vol. 10, no. 2, pp. 244–256, Apr. 2002.

[79] F. Pana and F. Put, "A survey on the evolution of rsvp," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 4, pp. 1859–1887, Fourth 2013.

[80] D. Vali *et al.*, "A survey of internet qos signaling," *IEEE Commun. Surveys Tuts.*, vol. 6, no. 4, pp. 32–43, Fourth 2004.

[81] S. Shenker, C. Partridge, and R. Guerin, "Specification of Guaranteed Quality of Service," RFC 2212 (Proposed Standard), Internet Engineering Task Force, Sep. 1997. [Online]. Available: http://www.ietf.org/rfc/rfc2212.txt

[82] K. Nichols *et al.*, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers," RFC 2474 (Proposed Standard), Internet Engineering Task Force, Dec. 1998, updated by RFCs 3168, 3260. [Online]. Available: http://www.ietf.org/rfc/rfc2474.txt

[83] J.-S. Li and C.-S. Mao, "Providing flow-based proportional differentiated services in class-based diffserv routers," *IEE Proceedings Communications*, vol. 151, no. 1, pp. 82–88, Feb 2004.

[84] R.-F. Liao and A. Campbell, "Dynamic core provisioning for quantitative differentiated services," *IEEE/ACM Trans. Netw.*, vol. 12, no. 3, pp. 429–442, June 2004.

[85] E. Rosen, A. Viswanathan, and R. Callon, "Multiprotocol Label Switching Architecture," RFC 3031 (Proposed Standard), Internet Engineering Task Force, Jan. 2001, updated by RFCs 6178, 6790. [Online]. Available: http://www.ietf.org/rfc/rfc3031.txt

[86] D. Awduche *et al.*, "Requirements for Traffic Engineering Over MPLS," RFC 2702 (Informational), Internet Engineering Task Force, Sep. 1999. [Online]. Available: http://www.ietf.org/rfc/rfc2702.txt

[87] ——, "RSVP-TE: Extensions to RSVP for LSP Tunnels," RFC 3209 (Proposed Standard), Internet Engineering Task Force, Dec. 2001, updated by RFCs 3936, 4420, 4874, 5151, 5420, 5711, 6780, 6790, 7274. [Online]. Available: http://www.ietf.org/rfc/rfc3209.txt

[88] B. Claise, "Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of IP Traffic Flow Information," RFC 5101 (Proposed Standard), Internet Engineering Task Force, Jan. 2008, obsoleted by RFC 7011. [Online]. Available: http://www.ietf.org/rfc/rfc5101.txt

[89] A. Farrel, J.-P. Vasseur, and J. Ash, "A Path Computation Element (PCE)-Based Architecture," RFC 4655 (Informational), Internet Engineering Task Force, Aug. 2006. [Online]. Available: http://www.ietf.org/rfc/rfc4655.txt

[90] A. Farrel and D. King, "Unanswered Questions in the Path Computation Element Architecture," RFC 7399 (Informational), Internet Engineering Task Force, Oct. 2014. [Online]. Available: http://www.ietf.org/rfc/rfc7399.txt

[91] V. Foteinos *et al.*, "Operator-friendly traffic engineering in ip/mpls core networks," *IEEE Trans. Netw. Service Manag.*, vol. 11, no. 3, pp. 333–349, Sept 2014.

[92] A. Elwalid *et al.*, "Mate: Mpls adaptive traffic engineering," in *Proc. 20th Annual Joint Conf. of the IEEE Comput. and Commun. Societies*, vol. 3, 2001, pp. 1300–1309 vol.3.

[93] S. Kandula *et al.*, "Dynamic load balancing without packet reordering," *SIGCOMM Comput. Commun. Rev.*, vol. 37, no. 2, pp. 51–62, Mar. 2007.

[94] J. Fernandez *et al.*, "Bandwidth aggregation-aware dynamic qos negotiation for real-time video streaming in next-generation wireless networks," *IEEE Trans. Multimedia*, vol. 11, no. 6, pp. 1082–1093, 2009.

[95] J. Brassil, "An empirical study of bonded heterogeneous wan access links," in *Proc. 14th IEEE Workshop on Local and Metropolitan Area Netw.*, 2005, pp. 6 pp.–6.

[96] S. Prabhavat *et al.*, "Effective delay-controlled load distribution over multipath networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 22, no. 10, pp. 1730–1741, 2011.

[97] ——, "On load distribution over multipath networks," *IEEE Commun. Surveys Tuts.*, vol. 14, no. 3, pp. 662 –680, quarter 2012.

[98] Y. Lee and Y. Choi, "An adaptive flow-level load control scheme for multipath forwarding," in *Networking ICN 2001*, ser. Lecture Notes in Computer Science, P. Lorenz, Ed. Springer Berlin Heidelberg, 2001, vol. 2093, pp. 771–779.

[99] K. Chebrolu, B. Raman, and R. R. Rao, "A network layer approach to enable tcp over multiple interfaces," *Wirel. Netw.*, vol. 11, no. 5, pp. 637–650, sep 2005.

[100] K. Chebrolu and R. Rao, "Bandwidth aggregation for real-time applications in heterogeneous wireless networks," *IEEE Trans. Mobile Comput.*, vol. 5, no. 4, pp. 388–403, 2006.

[101] *Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2 (Release 13)*, 3GPP TS 36.300, Rev. V13.2.0, 2015.

[102] *Technical Specification Group Services and System Aspects; General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access (Release 13)*, 3GPP TS 23.401, Rev. V13.0.0, 2014.

[103] *Technical Specification Group Services and System Aspects; Policy and charging control architecture (Release 13)*, 3GPP TS 23.203, Rev. V13.1.0, 2014.

[104] N. Ali, A. Taha, and H. Hassanein, "Quality of service in 3gpp r12 lte-advanced," *IEEE Commun. Mag*, vol. 51, no. 8, pp. 103–109, August 2013.

[105] H. Ekstrom, "Qos control in the 3gpp evolved packet system," *IEEE Commun. Mag*, vol. 47, no. 2, pp. 76–83, February 2009.

[106] *LTE; Evolved Universal Terrestrial Radio Access Network (E-UTRAN); S1 general aspects and principles Release 12)*, 3GPP TS 36.410, Rev. V12.0.0, 2014.

[107] *LTE; Evolved Universal Terrestrial Radio Access Network (E-UTRAN); S1 layer 1 (Release 13)*, 3GPP TS 36.411, Rev. V12.0.0, 2014.

[108] H. Raza, "A brief survey of radio access network backhaul evolution: part ii," *IEEE Commun. Mag*, vol. 51, no. 5, pp. 170–177, May 2013.

[109] "Backhaul Evolution - Integrated QoS Management," Deliverable, NGMN Alliance, December 2012. [Online]. Available: http://www.ngmn.org/uploads/media/NGMN_Backhaul_Evolution_-_Integrated_QoS_Management_Part_One_v1.3.pdf

[110] *Technical Specification Group Services and System Aspects; Architectural enhancements for end-to-end Quality of Service (QoS) (Release 7)*, 3GPP TR 23.802, 2005.

[111] *Technical Specification Group Services and System Aspects; Feasibility study on user plane congestion management (Release 12)*, 3GPP TR 22.805, Rev. V12.1.0, 2012.

[112] "Quality of Service in LTE," Whitepaper, Sandvine - Intelligent Broadband Networks, 2014. [Online]. Available: https://www.sandvine.com/downloads/general/whitepapers/quality-of-service-in-lte-long-form.pdf

[113] *DVB-RCS2 Higher Layer Satellite Specification*, ETSI TS 101 545-3, Rev. V1.2.1, April 2014.

[114] *Satellite Earth Stations and Systems (SES); Broadband Satellite Multimedia (BSM); QoS Functional Architecture*, ETSI TS 102 462, Rev. V1.1.1, December 2006.

[115] *Guidelines for the Implementation and Use of TS 101 545-3 (DVB-RCS2 Higher Layer)*, ETSI TR 101 545-5, Rev. V1.1.1, April 2014.

[116] M. Angeles Vazquez Castro and F. Vieira, "Dvb-s2 full cross-layer design for qos provision," *IEEE Commun. Mag*, vol. 50, no. 1, pp. 128–135, January 2012.

[117] "Software-Defined Networking: The New Norm for Networks," Whitepaper, Open Networking Foundation, Palo Alto, CA, USA, April 2012. [Online]. Available: http://www.opennetworking.org/images/stories/downloads/sdn-resources/white-papers/wp-sdn-newnorm.pdf

[118] B. A. A. Nunes *et al.*, "A survey of software-defined networking: Past, present, and future of programmable networks," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 3, pp. 1617–1634, Third 2014.

[119] W. Xia *et al.*, "A survey on software-defined networking," *IEEE Commun. Surveys Tuts.*, vol. PP, no. 99, pp. 1–1, 2014.

[120] *OpenFlow Switch Specification*, Open Networking Foundation Std., Rev. Version 1.3.2, April 2013.

[121] W. Wendong *et al.*, "Autonomic qos management mechanism in software defined network," *China Commun.*, vol. 11, no. 7, pp. 13–23, July 2014.

[122] H. Egilmez *et al.*, "Openqos: An openflow controller design for multimedia delivery with end-to-end quality of service over software-defined networks," in *Proc. Asia-Pacific Signal and Informat. Process. Assoc. Annu. Summit and Conf.*, Dec 2012, pp. 1–8.

[123] L. Bertaux *et al.*, "Software defined networking and virtualization for broadband satellite networks," *IEEE Commun. Mag*, vol. 53, no. 3, pp. 54–60, March 2015.

[124] "Service delivery over integrated satellite and terrestrial networks," Telecommunications and Integrated Application Study, European Space Agency, 2014. [Online]. Available: http://artes.esa.int/projects/service-delivery-over-integrated-satellite-and-terrestrial-networks

[125] A. Kapovits *et al.*, "Advanced topics in service delivery over integrated satellite terrestrial networks," in *Proc. 7th Adv. Satellite Multimedia Syst Conf. and 13th Signal Process. for Space Commun. Workshop*, Sept 2014, pp. 92–98.

[126] C. Niephaus, M. Kretschmer, and K. Jonas, "Qos-aware wireless back-haul network for rural areas in practice," in *Proc. Global Commun. Conf. Workshops*, Dec 2012, pp. 24–29.

[127] M. Kretschmer *et al.*, "Topology discovery and maintenance for heterogeneous wireless back-haul networks supporting unidirectional technologies," in *Proc. IEEE 10th Malaysia Int. Conf. on Communications*, Oct 2011, pp. 1–6.

[128] M. Kretschmer, C. Niephaus, and G. Ghinea, "Qos-aware flow monitoring and event creation in heterogeneous mpls-based wireless mesh networks supporting unidirectional links," in *Proc. IEEE 9th Malaysia Int. Conf. on Communications*, Dec 2009, pp. 680–685.

[129] M. Kretschmer, "Integration of unidirectional technologies into a wireless back-haul architecture," Ph.D. dissertation, School of Information Systems, Computing and Mathematics, Brunel University, 2012.

[130] *Methods for subjective determination of transmission quality*, ITU-T Recommendation P.800, 1996.

[131] K. Kilkki, "Quality of experience in communications ecosystem," *j-jucs*, vol. 14, no. 5, pp. 615–624, mar 2008.

[132] S. Chikkerur *et al.*, "Objective video quality assessment methods: A classification, review, and performance comparison," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 165–182, June 2011.

[133] S. Winkler and P. Mohandas, "The evolution of video quality measurement: From psnr to hybrid metrics," *IEEE Trans. Broadcast.*, vol. 54, no. 3, pp. 660–668, Sept 2008.

[134] T. Serif, S. R. Gulliver, and G. Ghinea, "Infotainment across access devices: The perceptual impact of multimedia qos," in *Proc. of the 2004 ACM Symp. on Appl. Comput.* ACM, 2004, pp. 1580–1585.

[135] T. Serif and G. Ghinea, "Mobile information access in the real world: A story of three wireless devices," *Computers in Human Behavior*, vol. 24, no. 4, pp. 1385 – 1403, 2008.

[136] M. Fiedler, T. Hossfeld, and P. Tran-Gia, "A generic quantitative relationship between quality of experience and quality of service," *IEEE Netw.*, vol. 24, no. 2, pp. 36–41, March 2010.

[137] G. Ghinea and J. Thomas, "An approach towards mapping quality of perception to quality of service in multimedia communications," in *Proc. IEEE 3rd Workshop on Multimedia Signal Process.*, Sept 1999, pp. 497–502.

[138] G. Ghinea, J. P. Thomas, and R. S. Fish, "Multimedia, network protocols and users - bridging the gap," in *Proc. of the 7th ACM Int. Conf. on Multimedia (Part 1)*, 1999, pp. 473–476.

[139] G. Ghinea and G. Magoulas, "Quality of service for perceptual considerations: an integrated perspective," in *Proc. IEEE Int. Conf. on Multimedia and Expo*, August 2001, pp. 571–574.

[140] S. Gulliver and G. Ghinea, "The perceptual and attentive impact of delay and jitter in multimedia delivery," *IEEE Trans. Broadcast.*, vol. 53, no. 2, pp. 449–458, June 2007.

[141] R. Stankiewicz and A. Jajszczyk, "A survey of qoe assurance in converged networks," *Comput. Netw.*, vol. 55, no. 7, pp. 1459–1473, May 2011.

[142] A. Callado *et al.*, "A survey on internet traffic identification," *IEEE Commun. Surveys Tuts.*, vol. 11, no. 3, pp. 37 –52, quarter 2009.

[143] K. Salehin and R. Rojas-Cessa, "Combined methodology for measurement of available bandwidth and link capacity in wired packet networks," *IET Commun.*, vol. 4, no. 2, pp. 240–252, January 2010.

[144] D. Pezaros, M. Hoerdt, and D. Hutchison, "Low-overhead end-to-end performance measurement for next generation networks," *IEEE Trans. Netw. Service Manag.*, vol. 8, no. 1, pp. 1–14, March 2011.

[145] F. Kuipers *et al.*, "Performance evaluation of constraint-based path selection algorithms," *IEEE Netw.*, vol. 18, no. 5, pp. 16–23, Sept 2004.

[146] "D1.1: Scenarios, requirements and kpis for 5g mobile and wireless system," Deliverable, METIS Project, April 2013. [Online]. Available: https://www.metis2020.com/wp-content/uploads/deliverables/METIS_D1.1_v1.pdf

[147] A. Meddeb, "Internet qos: Pieces of the puzzle," *IEEE Commun. Mag*, vol. 48, no. 1, pp. 86–94, 2010.

Christian Niephaus received his B.Sc. and M.Sc. in Computer Science from the University of Applied Sciences Bonn-Rhein-Sieg in 2006 and 2008, respectively. During his Master thesis he dealt with efficiency of MAC protocols for IEEE 802.11 networks. Since 2007 he is with the Fraunhofer Gesellschaft e.V where he participates in different EU FP6 and FP7 projects such as NetQOS, CARMEN, BATS as well as national research projects as contributor and task leader. He is also a PhD candidate at Brunel University London. His current research interests are emerging 5G networks and satellite-terrestrial network convergence with a strong focus on the special requirements of rural and remote areas. In this regard he (co-)authored a considerable number of papers in relevant international conferences, journals and book-chapters.

Mathias Kretschmer received the Dipl.-Ing. degree in communications engineering from the University of Applied Sciences Cologne, Cologne, Germany, and the Ph.D. degree in computer science from Brunel University, London, U.K. In 1998, he joined AT&T Labs Research, Florham Park, NJ, where he was involved in a number of multimedia distribution and streaming projects focusing on point-to-multipoint communication protocols and loss-resilient coding and transport. He then joined PreCache, a Sony Music-sponsored startup company to work on content-based routing and messaging. Since 2005, he has been with the Fraunhofer Institute for Open Communication Systems, Sankt Augustin, Germany, as a Senior Researcher working on next-generation wireless networks, where he also serves as the Deputy Department Head. He holds three patents in the field of communication and network protocols.

Dr. Gheorghita Ghinea is a Reader in the Computer Science Department at Brunel University, United Kingdom. He received the B.Sc. and B.Sc. (Hons) degrees in Computer Science and Mathematics, in 1993 and 1994, respectively, and the M.Sc. degree in Computer Science, in 1996, from the University of the Witwatersrand, Johannesburg, South Africa; he then received the Ph.D. degree in Computer Science from the University of Reading, United Kingdom, in 2000. His research activities lie at the confluence of Computer Science, Media and Psychology. In particular, his work focuses on the area of perceptual multimedia quality and building end-to-end communication systems incorporating user perceptual requirements. He has over 250 publications in leading international conferences and journals; Dr. Ghinea has co-edited two books on Digital Multimedia Perception and Design, and Multiple Sensorial Media Advance and Applications. He is a member of the IEEE and the British Computer Society.